Open Geospatial Data,
Software and Standards

ORIGINAL ARTICLE

Open Access

CrossMark

Remote sensing of burned areas via PCA, Part 2: SVD-based PCA using MODIS and Landsat data

Nikos Alexandris[1*], Nikos Koutsias[2] and Sandeep Gupta[3]

Abstract

Background: Singular value decomposition (SVD), as an alternative solution to principal components analysis (PCA), may enhance the spectral profile of burned areas in satellite image composites.

Methods: In this regard, we combine the pre-processing options of centering, non-centering, scaling, and non-scaling the input multi-spectral data, prior to the matrix decomposition, and treat their combinations as four different SVD-based PCA versions. Using both unitemporal and bi-temporal data sets, we test all four combinations to derive principal components. We assess the effects of the transformations based on multiresponse permutation procedures and quantify the enhanced spectral separability between burned areas and other major land cover classes via the Jeffries-Matusita metric. Lastly, we evaluate visually and numerically all principal components and select a subset of interest.

Results: The best transformation for the subset of selected components, is the uncentered-unscaled one.

Conclusions: The results indicate that an uncentered and unscaled SVD may improve the spectral separability of burned areas in some of the higher order components.

Keywords: PCA, EVD, SVD, Mean-centering, Scaling, Burned area mapping, MODIS, Landsat5 TM, Free open source software

Background

In the article *"Remote sensing of burned areas via PCA, Part 1: centering, scaling and EVD vs SVD."* [1], we present in-depth the concepts of PCA [2]; past scientific literature of PCA in remote sensing applications [3]; the link of PCA to burned area mapping [4]; the implications of centering and scaling [5]; and finally suggest that the uncentered-unscaled SVD-based PCA variant may further improve the spectral enhancement of burned area clusters compared to the conventional centered and EVD[1]-based PCA.

In multi-spectral imagery, burned areas build homogeneous clusters of low internal heterogeneity. Their mean spectral value is distanced from the composite's overall mean and they present lower projections, in some dimensions, in both uni- and multi-temporal composites. In the
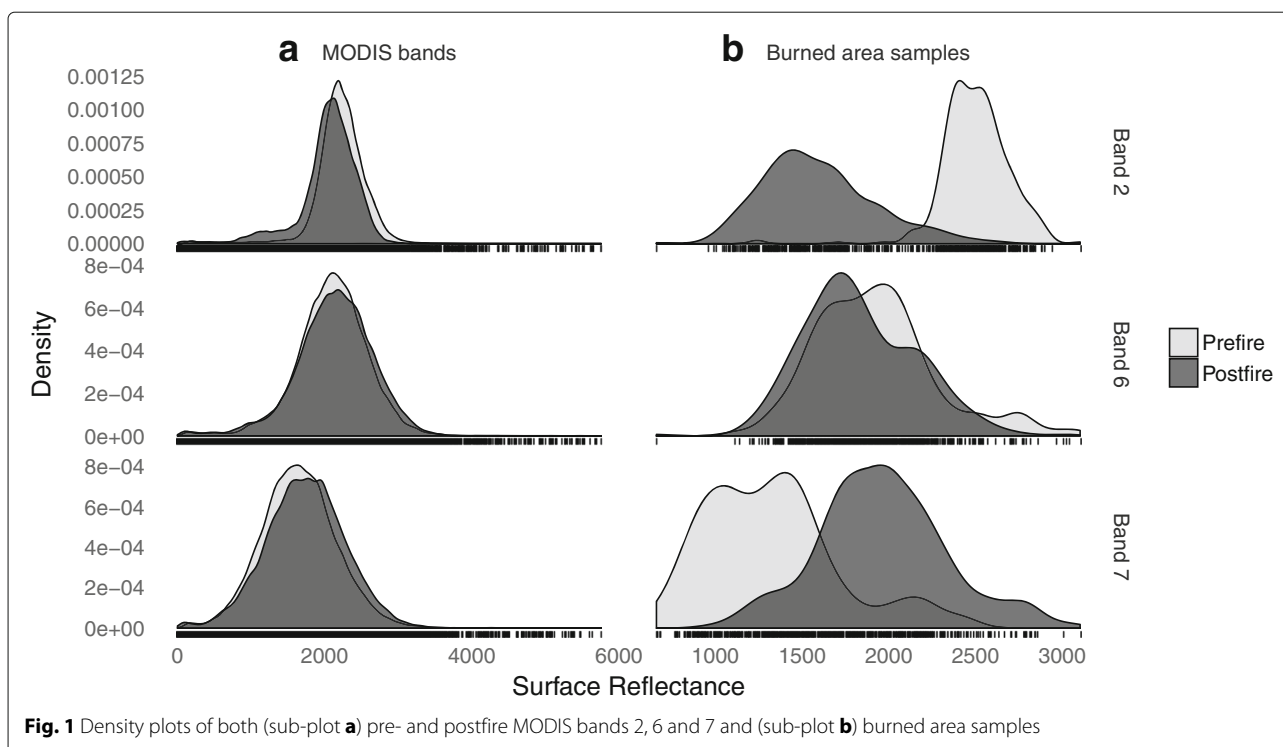
latter case, it is well noted that burned surfaces are absent in the prefire dimensions.

The pre-processing options to center and scale the image composites before the matrix decomposition, can be combined in different ways [2]. Their application influences the transformation of the spectral properties of burned area clusters. The impact of the transformations, is most evident in some of the higher order principal components. A non-centered SVD, captures in the first component greater amounts of information around the mean value of the input composite [5]. This can be advantageous in isolating burned clusters in some of the higher order components. Not scaling the input data may as well allow for subtle, yet useful, transformations applied in the initial dataset to be expressed in the restructured principal components. In this article, we demonstrate numerically the theoretical concepts of spectrally enhancing remotely sensed burned areas via SVD-based PCA. We apply and

*Correspondence: nik@nikosalexandris.net
[1]Independent Researcher, Ebringen, Germany
Full list of author information is available at the end of the article

Alexandris *et al. Open Geospatial Data, Software and Standards* (2017) 2:21

Page 2 of 26

Fig. 1 Density plots of both (sub-plot **a**) pre- and postfire MODIS bands 2, 6 and 7 and (sub-plot **b**) burned area samples
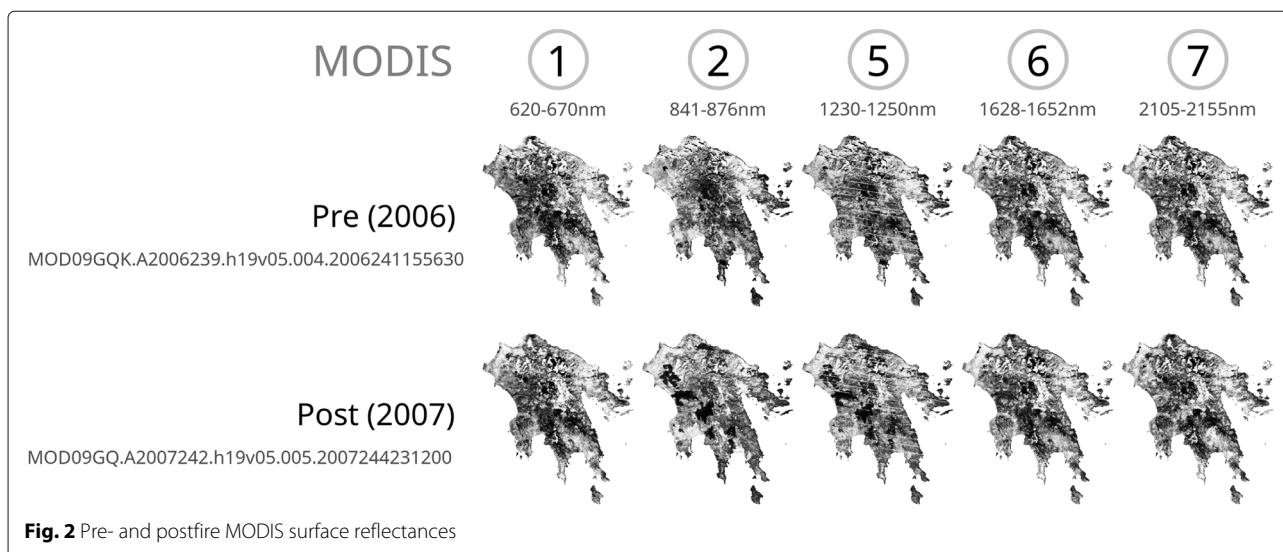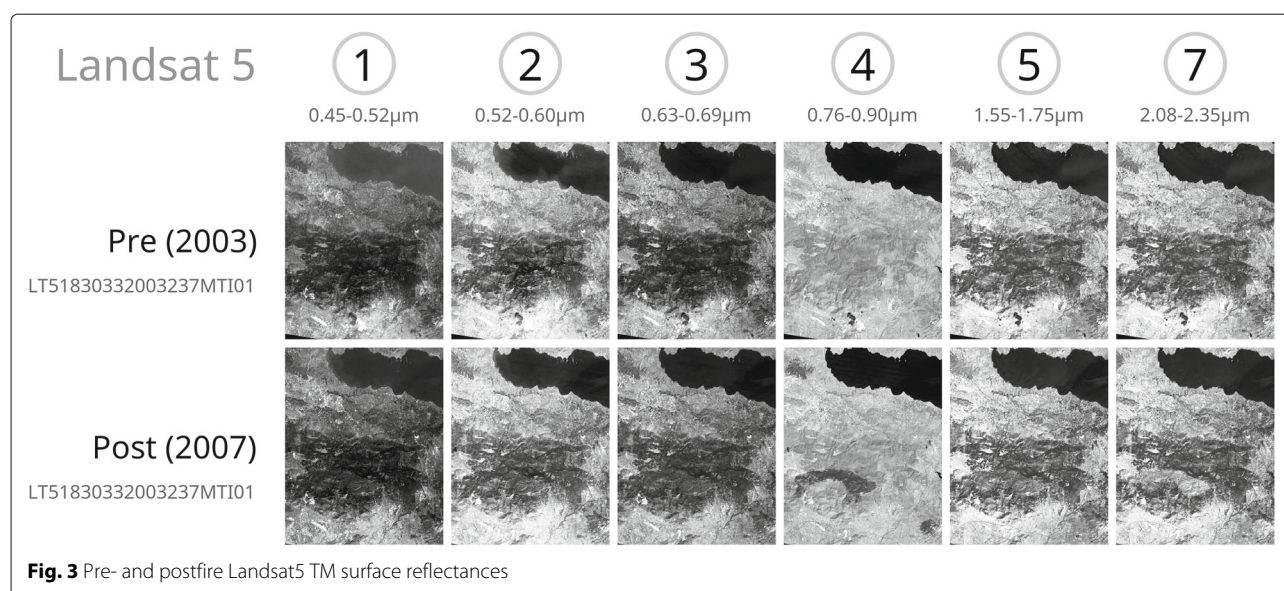
discuss the performance of four SVD versions. In addition, we go through an example-based quantitative discussion on the selection of the *best* principal components obtained via SVD.

Data

Within the first weeks, after the pause of large wildfires, burn scars absorbe higher amounts of solar energy. Compared to other surfaces, they present lower reflectance values in both Near-infrared (NIR) and Mid-

infrared (MIR) bands (Fig. 1) and appear expectedly darker than older burns. Therefore, post-fire multispectral imagery, needs to be timely acquired near after the pause of fires. Regarding pre-fire imagery in multitemporal data sets, they are best if acquired within the same season as the post-fire images. That is to hold the inter-seasonal reflectance variation of landscape features as low as possible. Generally, all scenes should be as cloudfree as possible, over large fire-affected regions in order to obtain more accurate results.

Fig. 2 Pre- and postfire MODIS surface reflectances

Fig. 3 Pre- and postfire Landsat5 TM surface reflectances

Based on the above, we analyse daily MODIS Terra L2G (MOD09GA)[2] and Landsat5 TM surface reflectance products (Figs. 2 and 3) respectively over Peloponnese and Mt Parnitha in Greece (Fig. 4). The selected MODIS acquisitions are a postfire scene in summer 2007 (Julian day 242)[3] and a prefire in summer 2006 (Julian day 239)[4]. MOD09 products are estimations of the surface spectral reflectance for each designated MODIS band and they are already atmospherically corrected. Variance-covariance and correlation coefficients for the selected input surface reflectance bands are presented in Table 1.

Worth mentioning is that MODIS band 5 (1.240 *Am*) is a very good discriminator with respect to the spectral response of burned areas (see sampled burned areas in Fig. 5 and refer to [6, 7]). However, in the acquired scene, band 5 is stripped, likely due to a calibration artefact causing anomalously high reflectance values [6]. Experimental transformations with data sets including band 5, derived noisy components. Therefore, this band has been excluded entirely from the analyses.

The Landsat5 TM scenes[5] were acquired in summer 2007 (Julian day 248, postfire)[6] and in summer 2003 (Julian day 237, prefire)[7]. These are already pre-processed data of Level-1[8] and delivered as scaled digital numbers. Since we do not cross-compare data from different sensors, and burned areas feature distinct spectral profiles, no further pre-processing was performed.

The selected MODIS scenes (Fig. 2) cover the Peloponnese peninsula (South Greece) with a total surface of $22,068\,km^2$ (main land of about $21,405\,km^2$ incl. surrounding islands on East, South). The Landsat5 TM products (Fig. 3) illustrate a region North of Athens–including Mt Parnitha–of about $1027\,km^2$. Both areas were severely damaged by large and uncontrolled wildland fires at the end of the summer 2007.

Tools

The employed methods were performed using free and open source software. Geospatial processing was performed using GRASS-GIS [8], QGIS [9] and FWTools [10]. The SVD-based PCA algorithm was applied via R's function *prcomp* [11]. Multi-Response Permutation
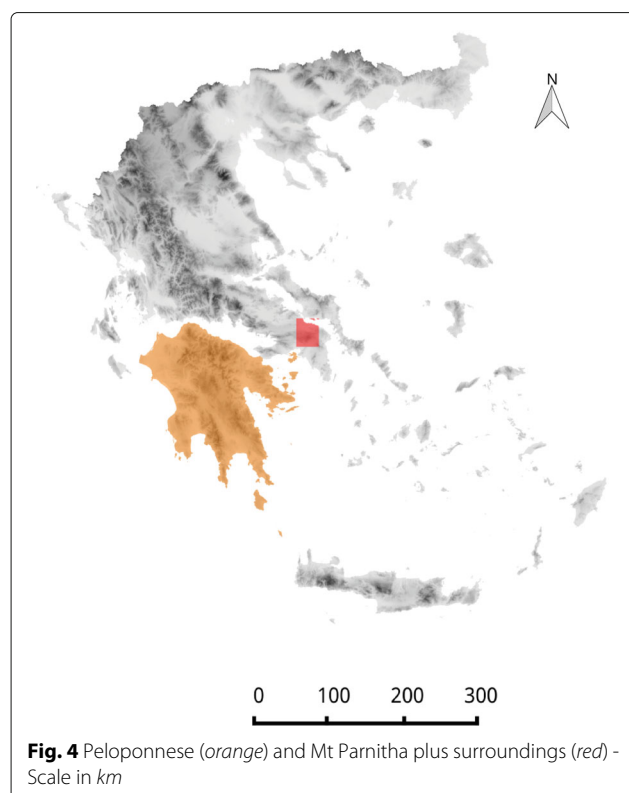

Fig. 4 Peloponnese (*orange*) and Mt Parnitha plus surroundings (*red*) - Scale in *km*

Alexandris *et al. Open Geospatial Data, Software and Standards* (2017) 2:21

Page 4 of 26

Table 1 Variance-covariance (white cells) and correlation coefficient (grey cells) matrices of MODIS and Landsat5 TM surface reflectance bands

Bands	1	2	6	7
1	0.0011	0.00099	0.002	0.0016
2	0.55	0.0028	0.0023	0.00098
6	0.92	0.67	0.0041	0.0031
7	0.90	0.34	0.89	0.0029

(a) Unitemporal composite, MODIS postfire bands 1, 2, 6 and 7

Bands	Pre 2	Pre 6	Pre 7	Post 2	Post 6	Post 7
Pre 2	0.0023	0.002	0.0014	0.0016	0.0016	0.0011
Pre 6	0.67	0.0037	0.003	0.0017	0.0033	0.0026
Pre 7	0.54	0.96	0.0027	0.0012	0.0027	0.0022
Post 2	0.63	0.52	0.45	0.0028	0.0023	0.00098
Post 6	0.51	0.84	0.8	0.67	0.0041	0.0031
Post 7	0.42	0.78	0.78	0.34	0.89	0.0029

(b) Bi-tempporal composite, MODIS pre- and postfire bands 2, 6 and 7

Bands	1	2	3	4	5	7
1	290	190	300	180	460	290
2	0.97	140	220	140	340	210
3	0.94	0.98	360	240	600	360
4	0.60	0.71	0.72	300	460	220
5	0.78	0.85	0.91	0.76	1200	700
7	0.82	0.86	0.91	0.61	0.95	440

(c) Unitemporal, Landsat5 TM postfire bands

Bands	Pre 2	Pre 4	Pre 7	Post 2	Post 4	Post 7
Pre 2	140.61	135.53	231.12	117.1	120.84	189.02
Pre 4	0.70	264.32	258.63	120.53	227.67	226.54
Pre 7	0.91	0.74	462.32	207.71	250.07	384.56
Post 2	0.85	0.64	0.83	135.19	144.29	211.04
Post 4	0.58	0.80	0.67	0.71	303.73	222.57
Post 7	0.76	0.66	0.85	0.86	0.61	443.63

(d) Bi-temporal composite, Landsat5 TM bands 2, 4 and 7

Procedures (MRPP) statistics were estimated using the *mrpp* and *meandist* functions, part of the R-package *vegan* [12]. The J-M index was implemented via custom R functions.

Methods

In the context of spectrally enhancing burned area clusters, we present uni- and bi-temporal study data sets. Therefore we label the four SVD-based PCA versions to derive principal components. Next, we describe the use of *multiresponse permutation procedures* to assess the effects of all transformations applied, namely centering, scaling and SVD itself. In addition, we refer to the Jeffries-Matusita spectral distance metric as a tool to quantify the separability between burned area and other major land cover class samples. Lastly, we overview an evaluation process for selecting principal components in which burned areas are spectrally enhanced. The complete workflow is visualised in Fig. 6.

Samples of burned areas and major land cover classes

Firstly, we delineated 42 samples of burned areas and numerous for vegetation and water bodies. Secondly, we extracted urban surfaces (greater than 200 ha) and bare ground samples from the CORINE 2000 land data map [13]. The samples, visualised in Fig. 7 are of both regular and irregular shape and consist by at least or more than 17 pixels[9]. We did avoid to digitise large and mixed samples that could result in high internal class heterogeneity.

Unitemporal and bitemporal composites
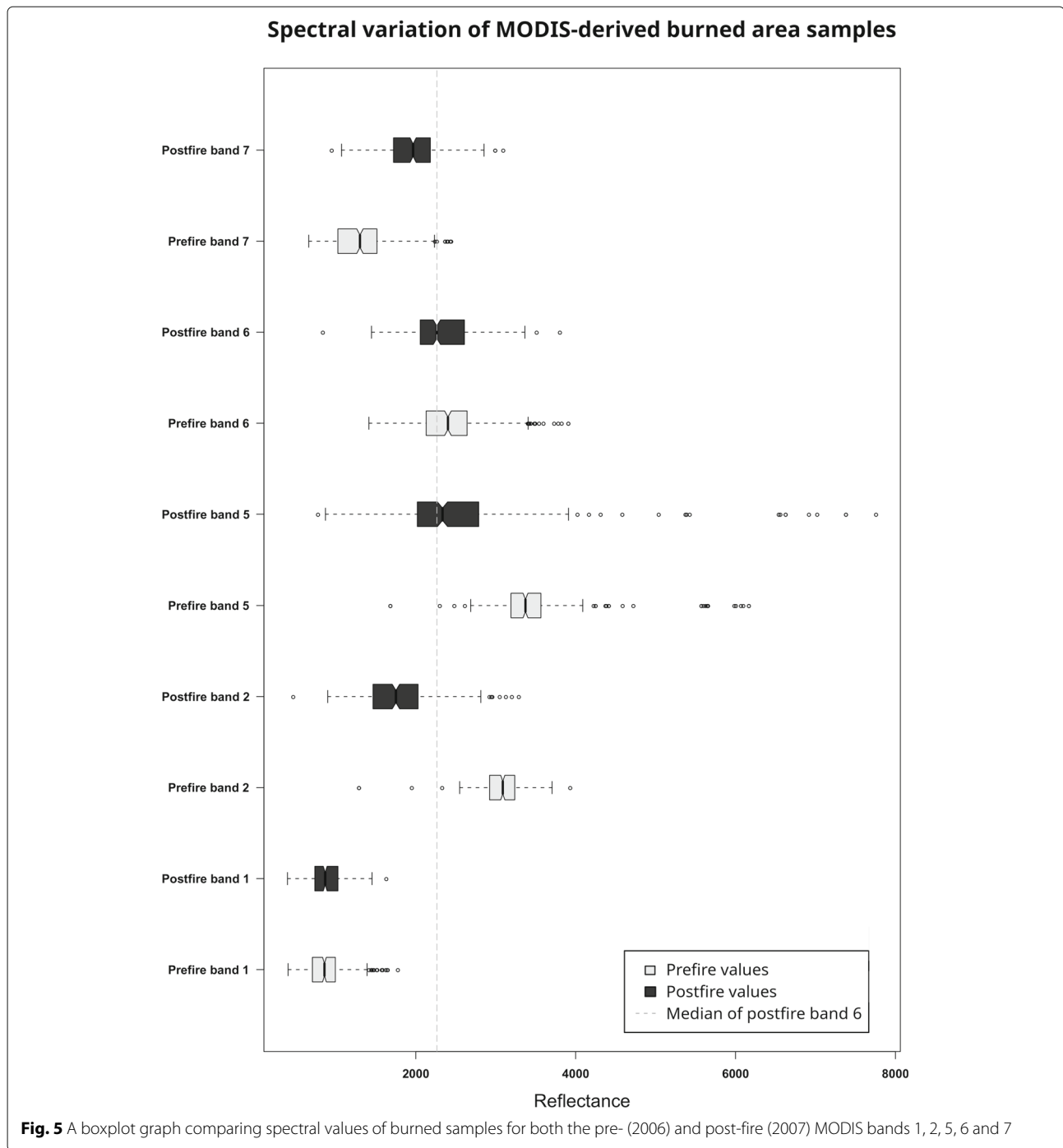
We define the following multi-spectral data sets:

1. Two unitemporal postfire sets: (a) a MODIS set build out of bands 1, 2, 6, 7 (in Fig. 2) and (b) a Landsat5 TM set composed of bands 1, 2, 3, 4, 5, 7 (in Fig. 3)
2. Two bi-temporal sets: (a) a MODIS composite build out of pre- and postfire bands 2, 6, 7 (in Fig. 2) and (b) a Landsat5 TM composite using pre- and postfire bands 2, 4, 7 (in Fig. 3)

The MODIS bands 1 and 2 were downscaled to 500 m to match the resolution of bands 6 and 7. The data sets will be cross-referenced as 1a, 1b, 2a and 2b hereafter. Scatterplot matrices for the samples in Fig. 7 extracted from both the unitemporal and bi-temporal MODIS composites are visualised in Figs. 12 and 13.

Four ways of extracting principal components via SVD

Employing SVD in burned area mapping applications, is an in-between enhancement step. It means to improve the performance of subsequent classification algorithms. Towards this end, we extract principal components via SVD from MODIS and Landsat5 TM surface reflectance data.

We subject to SVD the following versions of the data sets defined in the subsection "Unitemporal and bitemporal composites": (A) uncentered-unscaled, (B) uncentered-scaled, (C) centered-unscaled, (D) centered-scaled. Henceforth, the various versions will be referred as A, B, C and D respectively. Scatterplot matrices for the samples in Fig. 7 extracted from the MODIS-derived transformed images, are visualised in Figs. 14, 15, 16, 17, 18, 19, 20 and 21.

Alexandris *et al. Open Geospatial Data, Software and Standards* (2017) 2:21

Page 5 of 26

Fig. 5 A boxplot graph comparing spectral values of burned samples for both the pre- (2006) and post-fire (2007) MODIS bands 1, 2, 5, 6 and 7
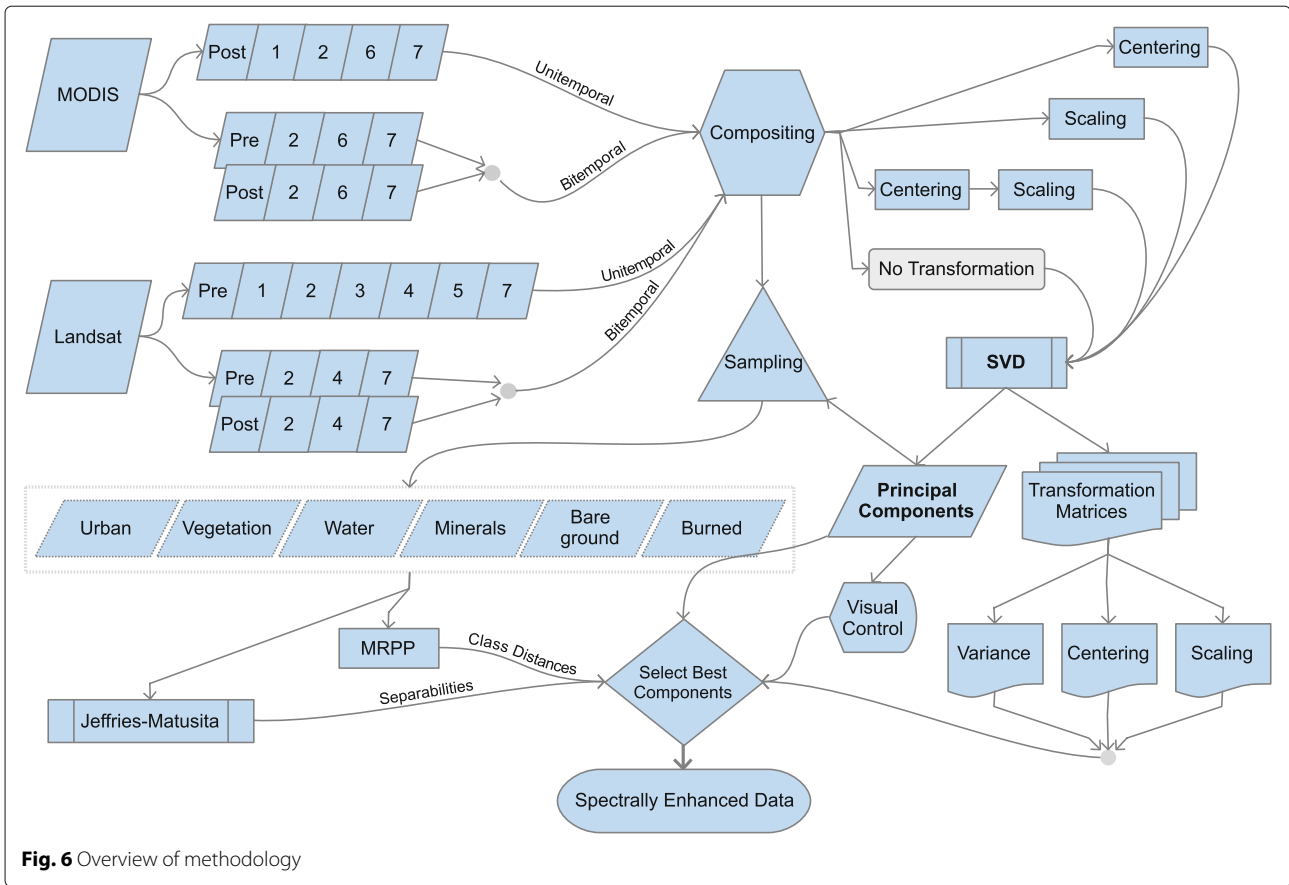
Multiresponse permutation procedures

Following multiresponse permutation procedures (MRPP) [14], one can describe the composition and configuration of major land cover class samples extracted from both the original and the transformed composites (Tables 2 and 3).

The MRPP null hypothesis (H_0) accepts no differences among the sampled classes.[10] This means that there is an equal chance for any possible combination of the data under H_0. The procedures estimate and compare the observed intra-class average distances (δ_o), weighted by their sample size (n), with average distances derived by all possible combinations ($\delta_{exp.}$) of the sampled data (permutations) expected under H_0. Essentially, they compare the dissimilarities within and among classes.

Alexandris *et al. Open Geospatial Data, Software and Standards*　(2017) 2:21

Page 6 of 26

Fig. 6 Overview of methodology

The significance of the test is reflected in the probability (*P*-value) of observing a mean distance δ as small or smaller than the observed δ_o under H_0. In addition, a measure of the within-class homogeneity is provided by $A = 1 - \delta_0/\delta_{exp.}$. The extreme case of all within-class observations being identical, equals to $\delta_o = 0$ and $A = 1$. Since the mean distance δ under H_0 is 0, an $A > 0$ represents within-class homogeneity and an $A < 0$ signifies within-class heterogeneity. Lastly, the classification strength [15] is the difference of the average between- and within-class dissimilarities.

The tests were performed using the complete set of observations sampled from the MODIS-based composites (in total 1085 pixels extracted from each band). However, due to the enormous amount of permutations demanded by the high number of observations sampled from Landsat5 TM data (in total 18865 pixels), we ran MRPP on 3000 randomly selected observations, independently for each Landsat5 TM-based data set. The euclidean distance metric was selected as the measure of dissimilarity between two observations.

Spectral distance metric

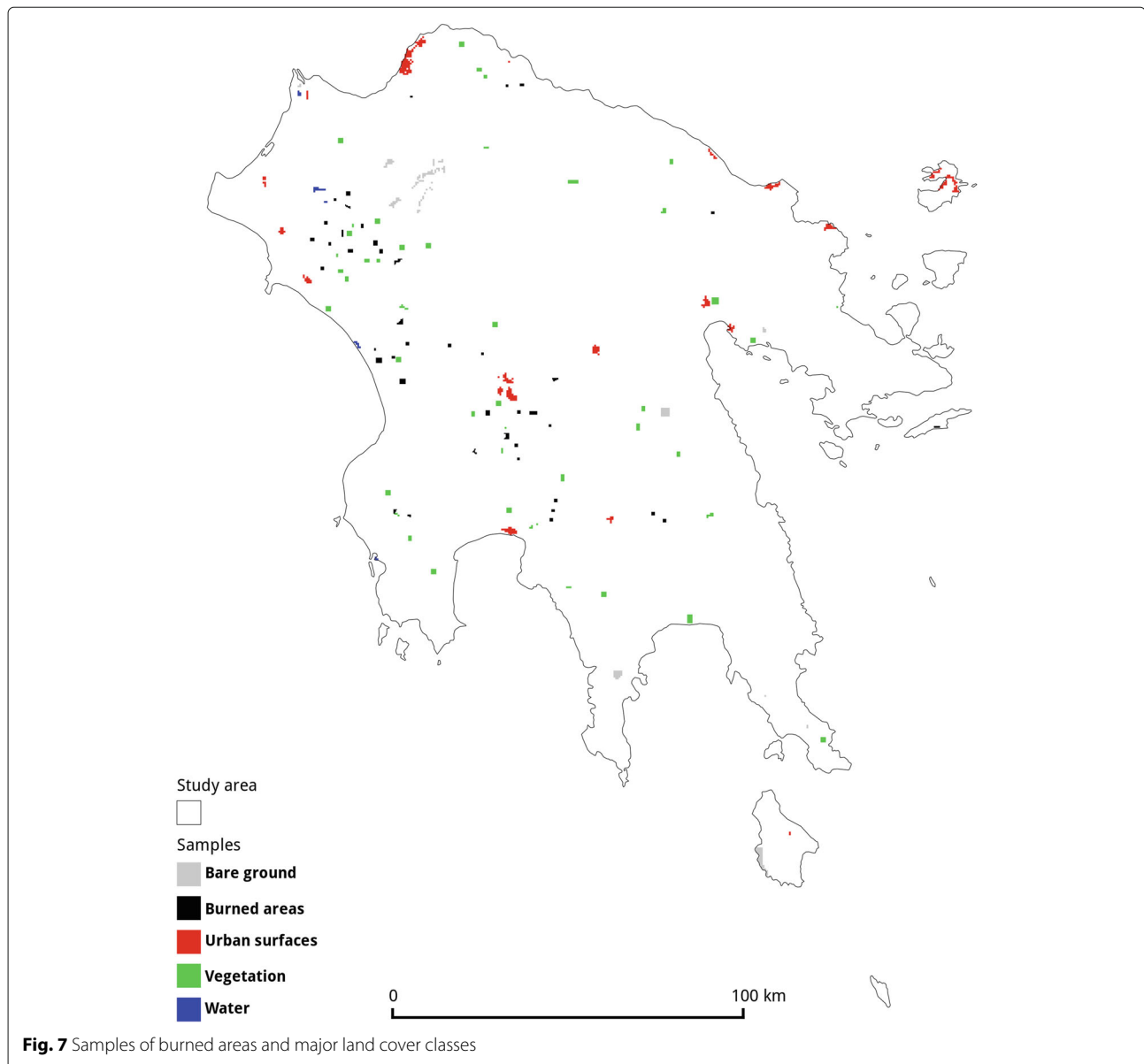The MRPP test assesses primarily the sampled burned area classe's quality of being different among the rest of the classes. Moreover, to verify numerically the effects of the pre-processing options mean-centering and scaling on the clusters of the sampled classes in terms of their configuration and composition. The procedures do not quantify, however, in a precise manner, the spectral enhancement of burned area samples after the application of SVD. To highlight how much the spectral separability, between burned and other class samples, increases or decreases, we rely on the Jeffries-Matusita (J–M) index.

J–M is well established in remote sensing applications as a measure of spectral separability between classes. The index is a transformation of the Bhatacharyya distance (Eq. 2) and applies to multivariate normal spectral class models. It is bound between $[0, 2.0]$ as defined by [16].

$$J_{ij} = 2 \left(1 - e^{-B}\right) \qquad (1)$$

where

$$B = 0.125 \left(i - j\right)^t \left\{\Sigma_i + \Sigma_j\right\}^{-1} \left(i + j\right)$$
$$+ 0.5 \, log_e \left\{ \frac{\left|\frac{(\Sigma_i + \Sigma_j)}{2}\right|}{\sqrt{|\Sigma_i| \, |\Sigma_j|}} \right\} \Rightarrow \qquad (2)$$

Alexandris *et al. Open Geospatial Data, Software and Standards* (2017) 2:21

Page 7 of 26

Fig. 7 Samples of burned areas and major land cover classes

where

B = Bhatacharyya index; i = first spectral signature vector; j = second spectral signature vector; Σ_i = covariance matrix of sample i; and Σ_j = covariance matrix of sample j.

Evaluation of the principal components

Selecting the components in which burn scars are emphasized, is important for any subsequent mapping attempt. The selection is rather a rejection scheme to filter out components that are dominated by information linked to unchanged landscape features. Likewise to reject ones that consist mainly of noise.

In this sense, we evaluate the outcomes of SVD considering in-depth the effects of the pre-processing

transformations centering and scaling via MRPP on samples of the land cover classes of interest; by visually inspecting the principal components; and comparing the *eigen*[11]*vectors*[12] and *eigen*[13]*values*[14].

Results and discussion

We discuss hereafter the results of the transformations and their impact on spatial distances within and between the sampled land cover classes. In addition, we compare the performance of the four SVD-based PCA versions in terms of the spectral enhancement of burned area clusters via the Jeffries-Matusita index. Next, we evaluate the principal components visually and numerically. Regarding the latter, we thoroughly review the case of the bi-temporal MODIS data set (2a), how its variance

Table 2 Statistics based on multiple response permutation procedures for MODIS and Landsat5 TM composites (based on *euclidean* distance, 999 permutations and significance for all deltas 0.001)

			MODIS			Landsat5 TM		
	Class	n	unitemporal 1a	bitemporal 2a	n	unitemporal 1b	n	bitemporal 2b
Distances δ	Urban areas	335	1004.0	1354.0	198	60.04	180	55.17
	Mineral extraction site				87	50.82	84	53.54
	Vegetation	307	815.9	1128.0	571	24.58	578	25.6
	Sparse vegetaion				1303	34.63	1296	26.41
	Bare ground	183	1361.0	1983.0	46	25.07	42	31.13
	Burned	188	796.8	1079.0	364	24.52	334	22.48
	Water	27	1231.0	1866.0	431	7.642	486	5.904
	A		0.4282	0.3851		0.619		0.6167
	Observed		979.8	1361.0		29.61		25.05
	Expected		1713.0	2214.0		77.72		65.35
Mean distances	Within classes		957.1141	1322.777		31.06279		24.56955
	Between classes		1971.5832	2404.687		94.58341		80.17993
	Overall		1713.4070	2129.348		77.72368		65.34757
	Classification strength		991.8072183	1058.7652870		64.9710787		55.1339036

is redistributed among the principal components. Finally, we justify the selection of the components that hold the highest separabilities.

Synopsis of pre-processing effects

Centering shifts the origin of the coordinate axes in the gravity center of the multidimensional data set. Scaling the centered dimensions forces unit variance before the analysis. In turn, this increases the influence of those variables with low variance and decreases the influence of those with high variance. Scaling, however, non-centered data does not yield to unit variance. It may even be mathematically questionable to do so, we do however include this combination for experimental completeness. While a centered SVD, equals the conventional EVD-based PCA, visual differences in terms of contrast may be perceived between components of the same order. These are atributed to the arbitrary sign in front of the eigenvectors.

Table 3 Statistics based on multiple response permutation procedures for principal components composites derived from MODIS composites (based on *euclidean* distance, 999 permutations and significance for all deltas 0.001)

			unitemporal 1a				bitemporal 2a			
			(A) Uncentered Unscaled	(B) Uncentered Scaled	(C) Centered Unscaled	(D) Centered Scaled	(A) Uncentered Unscaled	(B) Uncentered Scaled	(C) Centered Unscaled	(D) Centered Scaled
	Class	n								
Distances δ	Urban areas	335	1004.0	0.5189	1004.0	2.001	1354.0	0.5666	1354.0	2.502
	Vegetation	307	815.9	0.3931	815.9	1.576	1128.0	0.4756	1128.0	2.086
	Bare ground	183	1361.0	0.6867	1361.0	2.623	1983.0	0.8439	1983.0	3.589
	Burned	188	796.8	0.3823	796.8	1.518	1079.0	0.4701	1079.0	1.952
	Water	27	1231.0	0.5522	1231.0	2.245	1866.0	0.7531	1866.0	3.432
	A		0.4282	0.439	0.4282	0.4357	0.3851	0.4038	0.3851	0.3833
	Observed		979.8	0.4874	979.8	1.904	1361.0	0.5759	1361.0	2.495
	Expected		1713.0	0.8689	1713.0	3.374	2214.0	0.966	2214.0	4.046
Mean distances	Within classes		957.1141	0.4788438	957.1141	1.869949	1318.808	0.5572222	1318.808	2.423618
	Between classes		1971.5832	1.0020820	1971.5832	3.887467	2519.370	1.1055030	2519.370	4.599960
	Overall		1713.4070	0.8689210	1713.4070	3.374021	2213.834	0.9659689	2213.834	4.046094
	Classification strength		991.8072182	0.5146410	991.8072194	1.9835486	1158.0869603	0.5295941	1158.0869582	2.1047804

Alexandris *et al. Open Geospatial Data, Software and Standards* (2017) 2:21

Page 9 of 26

Within- vs between classes mean distances

We performed the MRPP test in order to diagnose the internal heterogeneity of burned area samples (within-class low dispersion of mean) and question their distinctness among other sampled land cover features (between-classes heterogeneity).

The within-classes heterogeneity is described by the *A* statistic and deviates little, in general, before and after the transformations–overall around 0.4 for MODIS data and around 0.6 for Landsat5 TM data. Hence, the transformations do not operate destructively in the internal structure of clusters for each class.

Before the transformations, the MRPP statistics show that burned area samples have relatively small mean within-class distance, which reflects their low within-class heterogeneity. For example, the respective δ values for burned area samples extracted from the MODIS data sets 1a and 2a, are 796.8 and 1079.0, lower than the observed δ_0 for all observations 979.8 and 1361.0 respectively (Table 2). In contrast, urban areas (and similarly mineral extraction sites in Landsat5 TM data) present similar δ than δ_0 values (i.e. 1004.0 and 1354.0 vs. 979.8 and 1361.0), yet higher than burned areas. Depending on the temporality of the samples extracted from the uni- or bi-temporal composites, burned area class distances δ are close to the ones of vegetation, sparse vegetation, and bare ground. A clear disjunction of water samples is present in all sampled data sets.

In the transformed data (Table 3), it is evident that centering does not alter the within- or between-classes spatial distances. The mean distances are identical for all MODIS-derived transformed composites (Table 3a, A and C of 1a and 1b) and practically equal for all Landsat5 TM-derived transformed composites (Table 4, A and C of 1b and 2b).

Scaling effects on both the range and the shape of the original point scatters are evident in the statistics (*A*, δ_0 and δ_{exp} values). For MODIS-based transformations,

nearly all scaled data sets result in higher *A* values than the unscaled data (Table 3 - 1a: 0.439 (B) vs. 0.4282 (A); 0.4357 (D) vs. 0.4282 (C); and Table 3-2a: 0.4038 (B) vs. 0.3851 (A)). An exception is the bitemporal centered-scaled data set which is practically the same as the centered data ((Table 3-2a: 0.3833 (D) vs. 0.3851 (C)). For the Landsat5 TM-based transformations, scaled bitemporal data have reduced *A* values while for the scaled unitemporal data they are close to the *A* values that correspond to the non-centered and centered data. Hence, low *A* and decreasing δ_o values, as observed for all scaled versions, reflect the suppresion of fine intra-class variations in the transformed data.

Lastly, we consider the classification strength values. Overall, the mean between-classes distances are higher than the within-classes distances for all data sets. For uncentered and centered data, both before and after the transformations, they are identical for the MODIS data sets (991.80 and 1058.76 in Table 2) and practically of equal importance for the Landsat5 TM data sets (64.97 and 55.13 in Table 2 and 67.35, 66.23 and 54.69, 54.62 in Table 4). In contrast, they are suppressed to low values and differ for all scaled versions. This translates in lower differences of within- and between-classes dissimilarities.

Estimation of class separabilities

Separability estimations between samples of burned areas and major land cover classes, quantify the magnitude of spectral enhancements. The indices are compared in a one-to-one manner, for all SVD versions, for each land cover class and principal component. Individual estimations and averages of the highest mean distances between samples of burned areas and major land cover classes can be extracted from Tables 5 and 6 for MODIS and Landsat data respectively.

In these tables, the row means correspond to the individual spectral separabilities between samples of burned areas, and other major land cover classes, for each

Table 4 Statistics based on multiple response permutation procedures for principal components composites derived from Landsat5 TM composites (based on *euclidean* distance, 999 permutations and significance for all deltas 0.001)

			unitemporal 1b								bitemporal 2b							
			(A) Uncentered Unscaled		(B) Uncentered Scaled		(C) Centered Unscaled		(D) Centered Scaled		(A) Uncentered Unscaled		(B) Uncentered Scaled		(C) Centered Unscaled		(D) Centered Scaled
	Class	*n*		*n*		*n*		*n*		*n*		*n*		*n*		*n*	
Distances δ	Urban areas	177	63.37	195	1.015	172	63.76	182	3.189	185	50.25	175	1.232	195	51.86	181	3.381
	Mineral extr. sites	96	55.29	94	0.829	109	53.0	98	2.458	106	55.34	92	1.076	83	61.1	88	3.147
	Vegetation	609	24.46	543	0.3469	570	24.82	569	1.2	526	25.63	569	0.4135	549	25.81	598	1.542
	Sparse vegetation	1309	33.74	1322	0.4945	1313	34.22	1262	1.537	1320	26.42	1314	0.5148	1298	26.63	1308	1.494
	Bare ground	33	32.57	50	0.441	46	29.88	38	1.464	50	30.43	42	0.6035	52	32.93	47	2.091
	Burned	309	26.69	326	0.3847	357	28.41	381	1.128	382	22.72	342	0.4425	346	22.46	338	1.25
	Water	467	7.612	470	0.1327	433	8.161	470	0.4872	431	5.842	466	0.1313	477	5.831	440	0.4605
	A		0.6258		0.6284		0.6167		0.6257		0.6116		0.5939		0.6119		0.5995
	Observed		29.49		0.4426		30.29		1.386		25.41		0.4481		25.4		1.496
	Expected		78.79		1.191		79.03		3.703		65.43		1.202		65.44		3.736
Mean distances	Within classes		30.17941		0.4479891		31.06418		1.392386		24.98853		0.4745462		24.84033		1.438458
	Between classes		96.84537		1.4637583		96.52606		4.507678		80.11150		1.4698200		80.02481		4.579308
	Overall		78.79091		1.1909819		79.03181		3.703097		65.43335		1.2018549		65.43857		3.735874
	Classification strength		67.3579873		1.0211541		66.2370470		3.1214813		54.6994877		0.9817204		54.6282804		3.0829241

Alexandris *et al. Open Geospatial Data, Software and Standards*   (2017) 2:21

Page 10 of 26

row-specific principal component. The column means correspond to the individual spectral separabilities between samples of burned areas and the column-specific land cover class for each version of SVD-based PCA. To exemplify, in Table 5, the average of the spectral separabilities between burned and other classes (first row) 0.959, 0.451, 1.178 and 1.005, extracted from principal component 1 derived from the uncentered and unscaled version of the unitemporal MODIS data set, is 0.898. The average of the spectral separabilities exclusively between samples of burned and urban areas (first column) 0.959, 0.122, 1.215 and 0.181, for components 1, 2, 3 and 4,

derived from the uncentered and unscaled version of the unitemporal MODIS data set, is 0.619.

Overall higher separabilities

For the unitemporal MODIS data set 1a, we gain higher overall average separabilities 0.722 in case uncentered-unscaled (A). The bi-temporal set 2a individuates the highest average 0.695 when the data are centered and scaled (D), practically identical to 0.694 when using uncentered-unscaled data (A). The corresponding average separation peaks for the Landsat5 TM sets, are 1.151 for the unitemporal set (1b) with uncentered-unscaled data

Table 5 Jeffries-Matusita matrix for burned area against major land cover class samples extracted from Principal Components derived from MODIS data sets

Data Set	Options	Principal Component	Urban	Vegetation	Bare ground	Water	Component means	Urban	Vegetation	Bare ground	Water	Component means
					Unscaled					Scaled		
Unitemporal PCs	Uncentered	1	0.959	0.451	1.178	1.005	0.898	0.98	0.009	1.113	0.999	0.775
		2	0.122	1.776	0.005	0.722	0.656	0.016	1.779	0.011	0.522	0.582
		3	1.215	0.305	0.873	1.082	0.869	1.244	0.546	0.97	1.102	0.966
		4	0.181	0.248	0.844	0.591	0.466	0.341	0.159	0.702	0.294	0.374
		Means	0.619	0.695	0.725	0.85	0.722	0.645	0.623	0.699	0.729	0.674
	Centered	1	0.884	0.081	1.125	1.058	0.787	0.983	0.03	1.131	0.994	0.784
		2	0.696	1.858	0.794	0.018	0.842	0.652	1.858	0.816	0.001	0.832
		3	0.891	0.044	0.209	0.727	0.468	0.981	0.028	0.361	1.246	0.654
		4	0.788	0.013	0.758	1.299	0.714	0.192	0.111	0.551	1.135	0.497
		Means	0.815	0.499	0.721	0.776	0.703	0.702	0.507	0.715	0.844	0.692
Bitemporal PCs	Uncentered	1	0.719	0.297	0.967	1.33	0.828	0.8	0.048	1.002	1.282	0.783
		2	0.23	1.325	0.707	0.114	0.594	0.209	1.43	0.544	0.165	0.587
		3	0.752	0.952	0.837	0.595	0.784	0.194	0.1	0.149	0.144	0.147
		4	1.422	1.221	1.461	1.213	1.329	1.678	1.556	1.745	1.392	1.593
		5	1.193	0.205	0.535	0.22	0.538	1.149	0.262	0.473	0.181	0.516
		6	0.062	0.087	0.079	0.145	0.093	0.076	0.087	0.037	0.196	0.099
		Means	0.73	0.681	0.764	0.603	0.694	0.684	0.58	0.658	0.56	0.62
	Centered	1	0.775	0.06	1.014	1.266	0.779	0.756	0.091	0.98	1.312	0.785
		2	0.121	1.667	0.018	0.587	0.598	0.078	1.585	0.018	0.754	0.609
		3	0.541	0.419	0.644	0.728	0.583	0.943	0.973	0.95	0.903	0.942
		4	1.245	0.804	1.375	1.55	1.244	1.201	0.79	1.275	1.414	1.17
		5	1.224	0.134	0.513	0.656	0.632	1.189	0.129	0.469	0.614	0.6
		6	0.06	0.094	0.077	0.034	0.066	0.061	0.092	0.073	0.038	0.066
		Means	0.661	0.53	0.607	0.803	0.65	0.705	0.61	0.627	0.839	0.695

Alexandris *et al. Open Geospatial Data, Software and Standards* (2017) 2:21

Page 11 of 26

Table 6 Jeffries-Matusita matrix for burned area against major land cover class samples extracted from Landsat5 TM data sets

Data Set	Principal Component	Urban	Minerals	Vegetation	Sparse vegetation	Bare ground	Water	Component means	Urban	Minerals	Vegetation	Sparse vegetation	Bare ground	Water	Component means
	Options	Unscaled							Scaled						
Unitemporal PCs — Uncentered	1	1.477	1.985	0.289	0.465	1.763	1.871	1.308	1.475	1.988	0.372	0.397	1.722	1.812	1.294
	2	0.366	0.726	1.973	0.436	0.591	1.998	1.015	0.309	0.666	1.997	1.249	0.049	1.998	1.045
	3	0.902	0.835	1.831	1.956	1.796	1.979	1.55	0.853	0.301	0.985	1.329	1.302	1.997	1.128
	4	1.627	1.99	1.977	1.603	1.606	1.709	1.752	1.672	1.993	1.993	1.908	1.865	1.866	1.883
	5	0.578	0.412	0.554	1.546	1.204	1.995	1.048	0.691	0.824	0.108	1.281	0.966	1.991	0.977
	6	0.498	0.277	0.102	0.181	0.076	0.243	0.23	0.488	0.286	0.113	0.222	0.09	0.232	0.238
	Means	0.908	1.037	1.121	1.031	1.173	1.633	1.151	0.915	1.01	0.928	1.064	0.999	1.649	1.094
Unitemporal PCs — Centered	1	1.414	1.98	0.527	0.336	1.739	1.928	1.321	1.507	1.99	0.273	0.474	1.73	1.68	1.276
	2	0.877	1.04	1.874	1.973	1.864	1.964	1.599	0.971	1.835	1.92	1.986	1.968	1.908	1.765
	3	1.577	1.868	1.998	1.781	0.871	1.999	1.682	1.446	1.511	1.993	1.49	0.318	1.999	1.46
	4	0.547	0.252	0.597	1.506	1.308	1.992	1.034	0.428	0.58	0.275	0.473	0.586	1.505	0.641
	5	0.465	1.168	0.011	0.48	0.496	0.753	0.562	0.687	1.07	0.413	1.455	0.754	1.961	1.057
	6	0.463	0.332	0.088	0.054	0.007	0.223	0.194	0.42	0.241	0.073	0.105	0.031	0.225	0.182
	Means	0.89	1.107	0.849	1.022	1.048	1.476	1.065	0.91	1.205	0.825	0.997	0.898	1.546	1.064
Bitemporal PCs — Uncentered	1	1.495	1.962	0.013	0.77	1.573	1.971	1.297	1.513	1.97	0.072	0.624	1.59	1.898	1.278
	2	0.674	1.492	1.938	0.535	0.974	1.439	1.175	0.59	1.395	1.953	0.5	0.971	1.659	1.178
	3	0.676	0.795	1.061	0.906	0.174	1.832	0.907	0.813	0.713	1.53	1.473	0.446	1.971	1.158
	4	1.854	1.979	1.992	1.999	1.994	2	1.97	0.742	0.381	0.462	0.755	1.005	1.996	0.89
	5	0.727	0.344	0.245	0.443	0.831	1.998	0.765	1.583	1.887	1.991	1.999	1.981	1.999	1.907
	6	0.615	0.188	0.077	0.11	0.083	0.025	0.183	0.499	0.241	0.056	0.217	0.286	0.178	0.246
	Means	1.007	1.127	0.888	0.794	0.938	1.544	1.05	0.957	1.098	1.011	0.928	1.046	1.617	1.109
Bitemporal PCs — Centered	1	1.459	1.96	0.076	0.58	1.588	1.965	1.271	1.536	1.97	0.005	0.726	1.571	1.883	1.282
	2	1.02	1.305	1.891	1.841	1.374	1.652	1.514	0.446	0.66	1.739	1.577	0.783	1.998	1.201
	3	0.679	0.793	1.088	0.926	0.178	1.83	0.916	0.733	1.023	0.283	0.038	0.413	0.305	0.466
	4	1.656	1.825	1.835	1.973	1.889	2	1.863	1.714	1.873	1.993	1.946	1.633	2	1.86
	5	0.891	0.424	0.018	0.859	1.105	0.851	0.691	0.624	0.344	0.572	1.72	1.622	0.527	0.902
	6	0.566	0.2	0.07	0.121	0.104	0.017	0.18	0.512	0.256	0.116	0.145	0.091	0.044	0.194
	Means	1.045	1.085	0.83	1.05	1.04	1.386	1.073	0.928	1.021	0.785	1.025	1.019	1.126	0.984

(A) and 1.109 for the bi-temporal set (2b) with uncentered but scaled data (B).

Cell-by-cell highest separabilities
Overall, when comparing the separability matrices in a cell-by-cell-manner (per class and component comparison), most of the highest observed values are concentrated in the uncentered-scaled case (B) followed by the uncentered-unscaled (A), leaving behind the other two cases. Cases A and C share most of the unitemporal-based highest separabilities, followed by the uncentered-scaled, leaving behind the centered-scaled data. For the bitemporal sets, uncentered-scaled (B) data count most of the highest separations followed by uncentered-unscaled, centered-scaled and lastly the centered-unscaled (C).

Per-component and per-class highest separabilities
Centered and scaled data (D) produce the highest separations in components 1 and 2 while uncentered-scaled data (B) attach to components 3, 4 and 6. The 5th component contains the smallest number of separation peaks, most of them contributed when using centered-unscaled data (C). Urban area samples are best separated from burned areas when using centered-unscaled data (C), while vegetated and bare ground samples with uncentered-unscaled data (D). Water surface samples peak their distance from burned areas twice in both uncentered-scaled (B) and centered-scaled (D) data. Mineral extraction sites peak once in uncentered-unscaled (A) and once in centered-scaled (D). Concluding, the most critical classes are best separated by using uncentered-unscaled data.

Visual inspection of the components
Visual inspection of the transformed images serves for quick control and is part of the complete evaluation process. On-sight, components 2, 3 and 4 are expected to be among the candidates in order to extract burned areas.

Alexandris *et al. Open Geospatial Data, Software and Standards* (2017) 2:21

Page 12 of 26

1. MODIS unitemporal data sets
   Burn scars are distinguished in all components derived from the unitemporal MODIS data set (1a, Fig. 8). For all SVD versions, burned areas appear very poor in the first component and rather blurry in the fourth component. Only the centered (both unscaled and scaled) second component represents sharply the scars. The uncentered components 2 and 3, appear to contain similar amounts of information linked to burned areas.

2. MODIS bi-temporal data set
   The bi-temporal MODIS composite (2a, Fig. 9) yields components in which we identify the burn scars within the 2nd, the 3rd and 4th components. The 3rd component appears occasionally unclear. Fragments of burn scars appear also in the 6th component, though they are rather noisy and stripped. In contrast, the 1st and the 5th components do not appear to hold distinguishable burned areas.

3. Landsat5 TM unitemporal data set
   On inspecting the components coming from the unitemporal postfire Landsat5 TM composite (1b, Fig. 10), the uncentered cases (A, B) distribute the scars on all components but the first. Also, they are barely visible in the 6th component. Conversely, in the centered but unscaled case (C) they appear more concentrated within the components 2, 3, 4 and noisy in components 5 and 6. Finally, the centered and scaled case (C) clearly displays the burnt signals in components 2 and 3 while the signal is rather weak in the remaining.

4. Landsat5 TM bi-temporal data set
   The outcomes based on the bi-temporal Landsat5 TM composite (2b, Fig. 11), include in all SVD versions a 2nd component that holds a moderate burnt signal. Component 3 is weaker for the uncentered cases and even more weak for the centered cases (C, D). Component 4 is best in cases
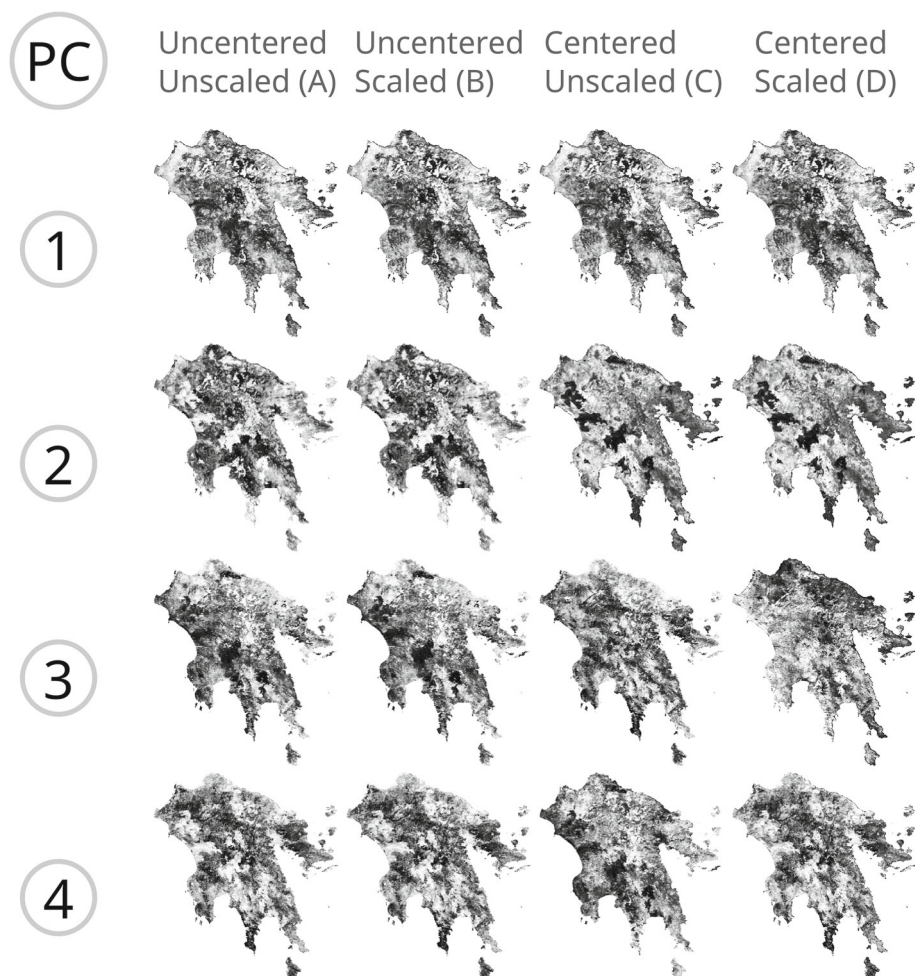

Fig. 8 Principal components derived from SVD of unitemporal MODIS composites (1a)
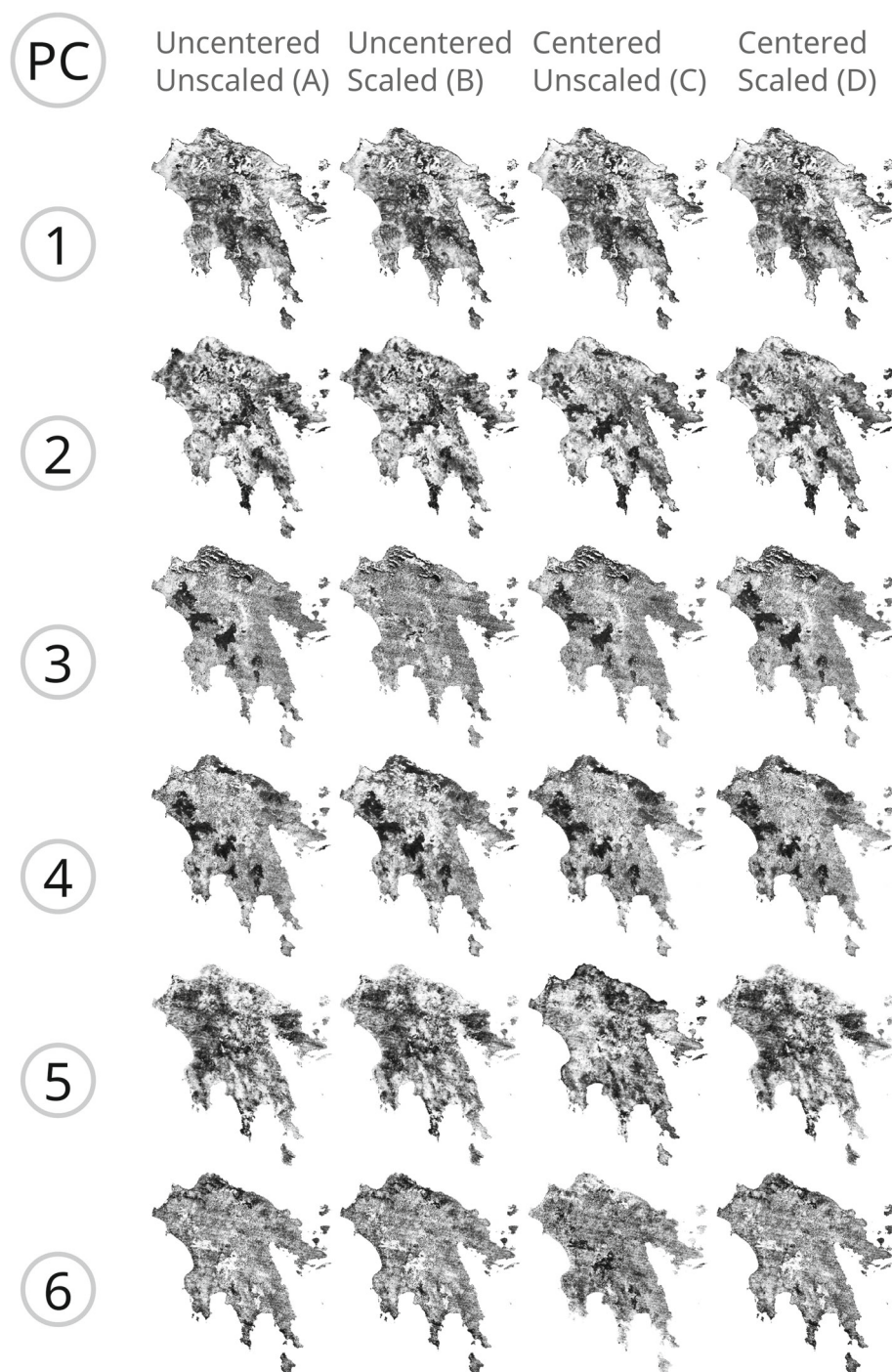
Fig. 9 Principal components as derived from SVD on the bi-temporal MODIS composite (2a)

A, C and D except for the case B where scars appear very weak if visible at all. The 5th component holds recognisable scars only in cases B and D.

Visually comparing the outcomes of the transformations allows for a rough similarity grouping of the images between *centered* and *uncentered*. As well, we observe, that the uncentered-scaled set of components, deviates from the uncentered-unscaled components.

Using the bi-temporal MODIS and the unitemporal Landsat5 TM composites, uncentered data highlight the burn scars in the third and fourth components while they appear weaker in the 2nd component (Figs. 9 and
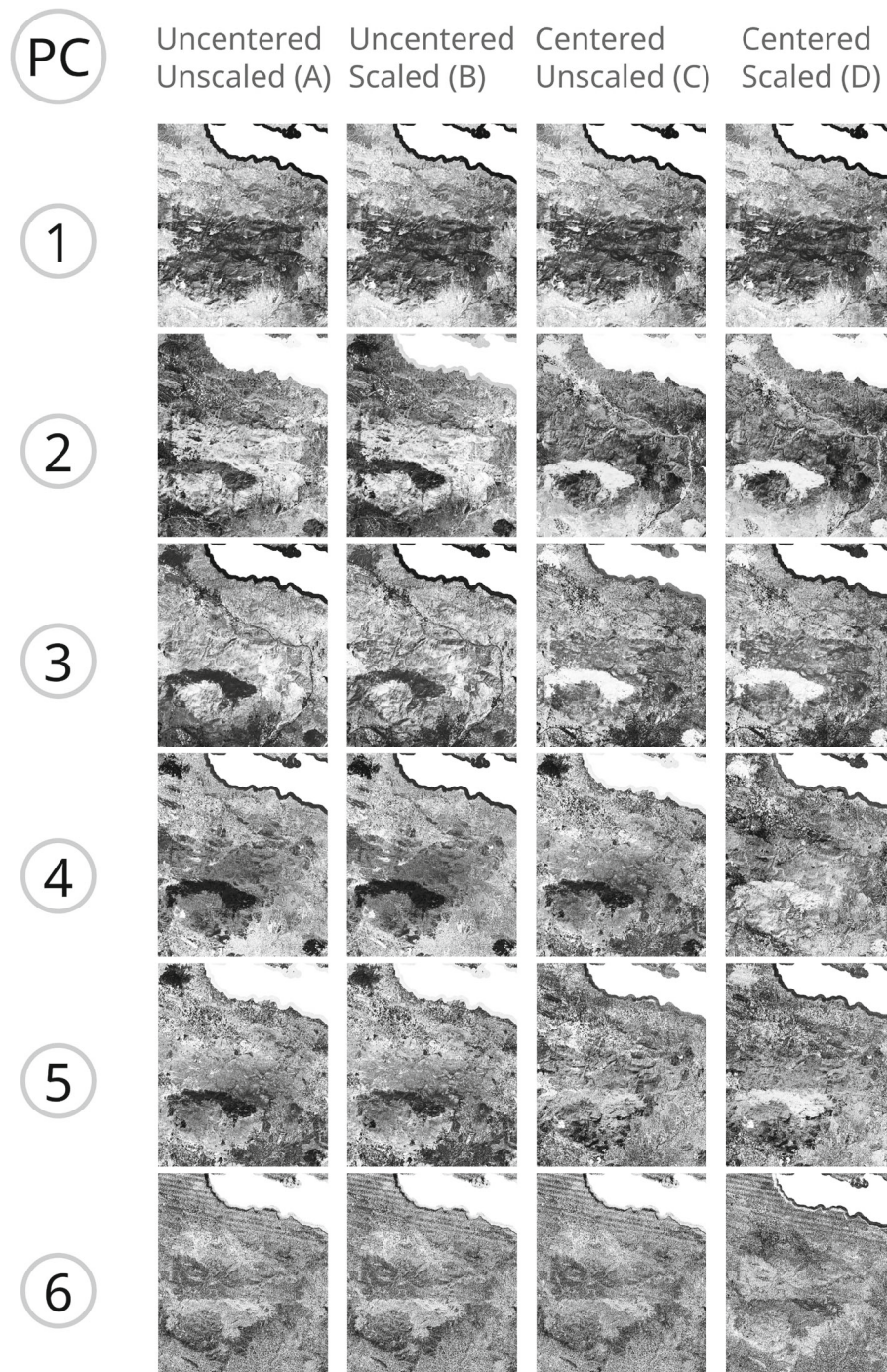
Alexandris *et al. Open Geospatial Data, Software and Standards*   (2017) 2:21

Page 14 of 26

Fig. 10 Principal components as derived from SVD on the unitemporal Landsat5 TM composite (1b)

10 respectively). Centered data emphasize the large burned surfaces within the second component and slightly alter their presence in the fourth component. An exception is the 4th centered-scaled transformed image, which seems very poor for the features of our interest.

Using the unitemporal MODIS data (1a), burn scars are divided among the second and third components. Finally, regarding the bi-temporal Landsat5 TM (2b) composite, uncentered-unscaled data spread the information in decreasing order of visual contrast against other features
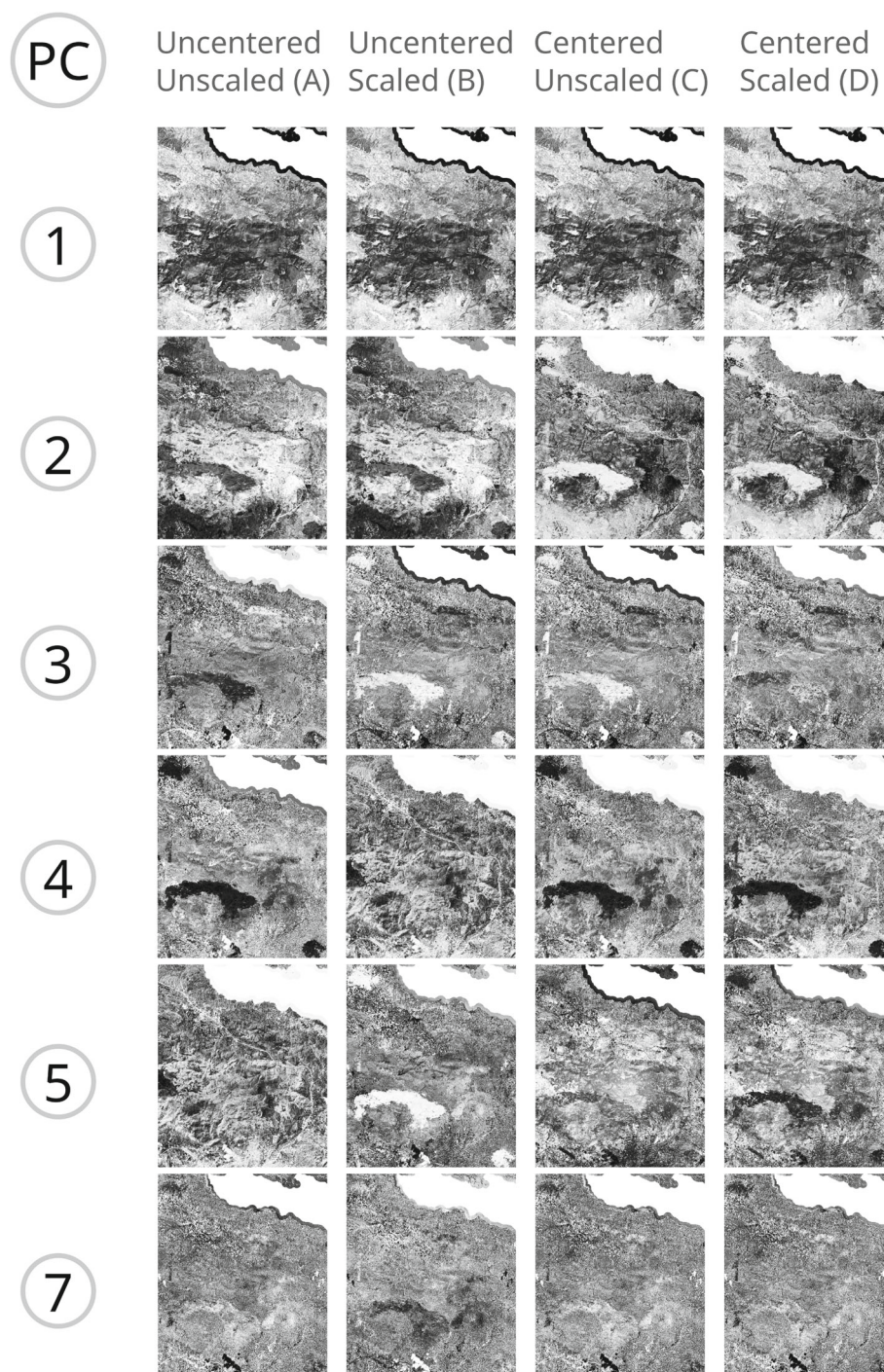
Alexandris *et al. Open Geospatial Data, Software and Standards*   (2017) 2:21

Page 15 of 26

Fig. 11 Principal components as derived from SVD on the bi-temporal Landsat5 TM composite (2b)

among the 4th, 2nd and 3rd components. The centered data, however, concentrate the scars in components 4 and 2 (Figs. 8 and 11).

Quantitative evaluation of the transformation matrices

Careful observation of the transformed variances expressed in percentage (%), reveals two groups of ranges for each component, depending on whether the input data matrix was centered or not (Table 7). This is expected as the first uncentered component passes through the origin of the coordinate system near to the centroid of the multidimensional point swarm. In the following sub-sections we discuss the effects of centering and scaling based on the transformation matrices derived

Table 7 Variance percentages form two groups for each principal component depending on whether centering is applied or not in the initial data set

		Uncentered		Centered	
		Unitemporal	Bi-temporal	Unitemporal	Bi-temporal
		Variances (%) *extracted from MODIS bands*			
Component	1	98.5 to 98.7	98.3 to 98.5	79.5 to 80	72.1 to 75.6
	2	1.2	0.9 to 1	17.5 to 17.8	13 to 14.9
	3	0.1 to 0.2	0.4	1.2 to 2	7.2 to 7.9
	4	0.1	0.2	0.9 to 1	3.6 to 4.2
	5	-	0	-	0.7 to 0.8
	6	-	0	-	0.2
		Variances (%) *extracted from Landsat5 TM bands*			
Component	1	98.09 to 98.14	97.34 to 97.35	85.77 to 88.60	79.70 to 78.90
	2	1.09 to 1.12	1.40 to 1.50	5.65 to 8.15	10.09 to 10.33
	3	0.53 to 0.57	0.46 to 0.56	4.86 to 5.12	4.71 to 5.02
	4	0.13 to 0.19	0.33 to 0.39	0.51 to 0.53	3.27 to 3.63
	5	0.05	0.26 to 0.27	0.31 to 0.33	1.66 to 1.86
	6	0 to 0.01	0.03 to 0.04	0.05 to 0.12	0.33 to 0.50

Table 8 Transformation matrices derived from SVD on the bi-temporal MODIS composite (corresponding components in Fig. 9)

Acquisition	Bands		Singular Value Decomposition											
			Uncentered Unscaled (A)						Uncentered Scaled (B)					
			1	2	3	4	5	6	1	2	3	4	5	6
Prefire (2006)	2	Eigenvectors	-0.48	0.48	0.52	-0.45	0.03	0.26	-0.41	0.48	0.18	-0.69	0.07	0.31
	6		-0.45	-0.24	0.37	0.32	-0.48	-0.52	-0.41	-0.08	0.36	-0.04	-0.61	-0.56
	7		-0.28	-0.37	0.28	0.50	0.54	0.40	-0.41	-0.39	0.60	0.28	0.41	0.27
Postfire (2007)	2		-0.45	0.53	-0.50	0.33	0.29	-0.28	-0.41	0.59	-0.21	0.49	0.33	-0.31
	6		-0.45	-0.26	-0.49	-0.10	-0.45	0.52	-0.41	-0.08	-0.39	0.29	-0.49	0.58
	7		-0.29	-0.47	-0.16	-0.58	0.42	-0.39	-0.41	-0.50	-0.54	-0.34	0.31	-0.28
	Standard deviation		6098.90	582.07	376.46	277.50	118.16	59.05	2.43	0.25	0.16	0.12	0.05	0.02
	Variance (%)		98.5	0.9	0.4	0.2	0	0	98.3	1	0.4	0.2	0	0
			Centered Unscaled (C)						Centered Scaled (D)					
			1	2	3	4	5	6	1	2	3	4	5	6
Prefire (2006)	2	Eigenvectors	0.28	0.48	0.56	-0.57	0.00	-0.24	0.35	0.54	0.60	-0.43	0.01	0.20
	6		0.49	-0.11	0.38	0.30	0.47	0.53	0.46	-0.12	0.27	0.33	-0.53	-0.55
	7		0.40	-0.23	0.30	0.47	-0.55	-0.42	0.44	-0.26	0.24	0.54	0.51	0.37
Postfire (2007)	2		0.30	0.73	-0.40	0.25	-0.29	0.28	0.33	0.65	-0.54	0.21	0.25	-0.25
	6		0.52	-0.08	-0.49	-0.08	0.46	-0.51	0.45	-0.12	-0.43	-0.15	-0.49	0.57
	7		0.39	-0.42	-0.22	-0.54	-0.42	0.39	0.41	-0.43	-0.16	-0.59	0.38	-0.36
	Standard deviation		1176.77	491.44	373.39	264.85	116.92	58.05	2.08	0.94	0.69	0.50	0.20	0.10
	Variance (%)		74.7	13.0	7.5	3.8	0.7	0.2	72.1	14.9	7.9	4.2	0.7	0.2

from SVD on the bi-temporal MODIS composite 2a (Fig. 9). All numbers compared beloware drawn from Table 8. The transformations matrices for composites 1a, 1b and 2b are presented in Tables 9, 10 and 11.

A subtlety that affects the numerical accuracy of calculations is the divisor N used for the covariance matrix in the *princomp* function (an EVD-based PCA implementation) and the divisor $N-1$ used in the *prcomp* function (an SVD implementation) [17]. Though this should practically make no difference for samples containing more than 30 observations.

Variance

In general, uncentered data practically channel all of the original's data variance in the 1st component (variances 98.5% and 98.3% for cases A, B respectively). On the other hand, centered data distribute significant amounts of information in higher order components (variances 74.7%, 72.1%, 74.7%, 72.1%, for cases C, D respectively).

For all cases, the variances of the last components (5th and 6th) are very low, while, as expected, the highest ones are identified in the major component (1st). In general, one can safely ignore these components since the former can be attributed to residual information and the latter mainly to unchanged features. Thus, we focus on the 2nd, 3rd and 4th components. The distribution of each original band in the transformed images is reflected in the eigen vectors, which act as weighting coefficients.

Centering

Centering decreases the absolute standard deviations of the extracted components. Yet the variance percentages of the higher order components increase substantially. This signifies that important amounts of the initial variation are redistributed among the higher order components 2, 3 and 4. On the contrary, performing the analysis without centering results in higher absolute standard deviations. Nonetheless, the variance percentages of the higher order components are substantially reduced in comparison to the 1st component. We then observe that centering relocates a lot of the information included in postfire band 2 in the 2nd component (eigenvector increases from 0.53 in case A to 0.73 in case C).

Burned surfaces are recorded as lower reflectance values in most of the spectral bands. Assuming they form data clusters which are clearly separated from the mean, the biggest portion of spectral information channeled in the 1st uncentered component, resembles mostly features other than burned. Postfire band 7-sourced information, increases in the 1st and 3rd centered components (respectively from 0.29 and 0.16 to 0.39 and 0.22) and decreases in the 2nd and 4th components (from 0.47 and

Table 9 Transformation matrices derived from SVD on unitemporal MODIS composites (corresponding components in Fig. 8)

Acquisition	Bands		Singular Value Decomposition							
			Uncentered Unscaled (A)				Uncentered Scaled (B)			
			1	2	3	4	1	2	3	4
Postfire (2007)	1	Singular vectors	0.25	0.24	0.94	0.05	0.5	0.31	0.81	0.07
	2		0.61	-0.73	0.01	0.30	0.50	-0.80	-0.03	0.34
	6		0.63	0.23	-0.19	-0.72	0.50	-0.04	-0.23	-0.83
	7		0.41	0.60	-0.30	0.63	0.50	0.52	-0.55	0.43
	Standard deviation		4421.23	490.42	116.73	114.29	1.99	0.22	0.09	0.05
	Variance (%)		98.65	1.21	0.07	0.07	98.53	1.21	0.21	0.05
			Centered Unscaled (C)				Centered Scaled (D)			
			1	2	3	4	1	2	3	4
Postfire (2007)	1	Singular vectors	0.34	-0.12	0.66	0.66	0.54	-0.17	-0.82	-0.05
	2		0.4	0.85	0.22	-0.27	0.38	0.87	0.05	0.32
	6		0.68	-0.05	-0.67	0.3	0.55	0	0.41	-0.73
	7		0.51	-0.51	0.27	-0.63	0.51	-0.47	0.39	0.61
	Standard deviation		936.95	442.49	115.26	102.25	1.78	0.84	0.28	0.18
	Variance (%)		79.99	17.84	1.21	0.95	79.54	17.57	2.03	0.85

Table 10 Transformation matrices derived from SVD on the unitemporal Landsat5 TM composite (corresponding components in Fig. 10)

Acquisition	Bands		Singular Value Decomposition											
			Uncentered Unscaled (A)						Uncentered Scaled (B)					
			1	2	3	4	5	6	1	2	3	4	5	6
Postfire (2007)	1	Eigenvectors	0.48	0.53	-0.52	-0.38	0.11	0.24	0.41	0.34	-0.5	-0.51	0.19	0.42
	2		0.25	0.14	-0.21	0.26	0.02	-0.9	0.41	0.16	-0.37	0.13	-0.05	-0.81
	3		0.3	-0.12	-0.25	0.83	0.13	0.36	0.41	-0.21	-0.24	0.76	0.03	0.4
	4		0.4	0.47	0.69	0.16	-0.33	0.06	0.4	0.6	0.58	0.1	-0.34	0.09
	5		0.59	-0.52	0.29	-0.26	0.47	-0.05	0.41	-0.27	0.45	-0.12	0.72	-0.12
	7		0.31	-0.44	-0.23	-0.13	-0.8	0.02	0.41	-0.61	0.08	-0.36	-0.57	0.03
	Standard deviation		162.84	17.20	12.45	6.02	3.79	1.15	2.43	0.26	0.18	0.11	0.06	0.03
	Variance (%)		98.14	1.09	0.57	0.13	0.05	0.00	98.09	1.12	0.53	0.19	0.05	0.01

Acquisition	Bands		Centered Unscaled (C)						Centered Scaled (D)					
			1	2	3	4	5	6	1	2	3	4	5	6
Postfire (2007)	1	Eigenvectors	0.3	0.49	-0.47	0.01	-0.61	0.29	0.41	0.37	-0.46	0.42	-0.45	-0.34
	2		0.22	0.2	-0.3	0.04	0.07	-0.9	0.43	0.16	-0.35	-0.13	0.17	0.79
	3		0.37	0.24	-0.29	0.27	0.75	0.31	0.43	0.12	-0.1	-0.66	0.34	-0.49
	4		0.27	-0.73	-0.53	-0.33	-0.01	0.07	0.34	-0.87	-0.22	0.22	0.11	-0.09
	5		0.69	-0.26	0.45	0.44	-0.23	-0.04	0.42	-0.14	0.54	-0.27	-0.65	0.14
	7		0.41	0.26	0.35	-0.79	0.13	0.01	0.41	0.21	0.56	0.5	0.47	-0.04
	Standard deviation		49.29	12.45	11.55	3.81	2.91	1.13	2.27	0.70	0.55	0.17	0.14	0.09
	Variance (%)		88.60	5.65	4.86	0.53	0.31	0.05	85.77	8.15	5.12	0.51	0.33	0.12

0.58 to 0.42 and 0.54) which might be also interpreted as a loss of useful information from the higher order components 2 and 4.

Scaling
While the effect of centering is obvious in both the eigen values (or singular) and vectors, scaling the input data deals with finer details. Depending on whether the dimensions to be scaled are already centered or not, the influence on the variance percentages of the extracted components varies. The variance changes very little, and only for the first two components, when using uncentered input data. Quite the opposite, using centered input data produces different percentages.

In general, scaling reduces the variance of the 1st component. The variance percentages for component 2 increase from 0.9 to 1% and 13 to 14.9% in cases A, B respectively. In the higher order components 3 and 4, scaling of the uncentered input data does not alter the variance percentages 0.4% and 0.2% respectively for cases A, B. The same is observed when using centered input data sets with respect to components 5 and 6 whose variances are 0.7% and 0.2% for cases C, D. This does not hold true, however, for components 3 and 4 where the numbers increase: 7.5 to 7.9% and 3.8 to 4.2% for cases C, D.

Worth emphasising is that scaling uncentered data prior to SVD relocates the biggest proportion of information originating from both the prefire and the postfire band 2 in components other than the 3rd. For case B, the prefire band 2 loadings in the 3rd component decrease from 0.52 to 0.18. Most of the prefire band 2 information is clearly channeled in the 4rth component (loading −0.69). The postfire band 2 loading in the 3rd component decreases as well from 0.50 to 0.21. Thus, burned areas appear isolated in the 4rth component (Fig. 9).

Selecting components with highest separabilities
Most of the highest per-class separability peaks, exist within the uncentered-unscaled data followed by the uncentered-scaled, the centered-scaled, and lastly the centered-unscaled data set. Yet, observations of the highest mean separabilities only, whether per-SVD version or per-class, do not suffice for selecting the best

Alexandris *et al. Open Geospatial Data, Software and Standards* (2017) 2:21

Page 19 of 26

Table 11 Transformation matrices derived from SVD on the bi-temporal Landsat5 TM composite (corresponding components in Fig. 11)

Acquisition	Bands		Singular Value Decomposition											
			Uncentered Unscaled (A)						Uncentered Scaled (B)					
			1	2	3	4	5	6	1	2	3	4	5	6
Prefire (2003)	2		0.32	-0.08	0.41	-0.06	0.6	-0.6	0.41	0.03	-0.46	0.52	0.25	0.54
	4	Eigenvectors	0.52	0.42	0.23	-0.63	-0.22	0.23	0.41	0.45	-0.06	-0.3	0.62	-0.39
	7		0.37	-0.52	0.52	0.36	-0.39	0.2	0.41	-0.5	-0.54	-0.42	-0.28	-0.21
Postfire (2007)	2		0.31	-0.09	-0.19	0.15	0.63	0.66	0.41	0.02	0.21	0.6	-0.33	-0.56
	4		0.5	0.48	-0.31	0.58	-0.16	-0.24	0.41	0.52	0.2	-0.31	-0.54	0.38
	7		0.38	-0.56	-0.61	-0.32	-0.12	-0.23	0.41	-0.53	0.64	-0.11	0.27	0.25
	Standard deviation		132.12	16.75	9.08	7.71	6.97	2.42	2.42	0.29	0.18	0.15	0.12	0.05
	Variance (%)		97.34	1.56	0.46	0.33	0.27	0.03	97.35	1.40	0.56	0.39	0.26	0.04
			Centered Unscaled (C)						Centered Scaled (D)					
			1	2	3	4	5	6	1	2	3	4	5	6
Prefire (2003)	2		0.28	0.14	-0.41	0.17	-0.56	-0.63	0.42	0.31	-0.47	0.43	-0.19	-0.53
	4	Eigenvectors	0.37	-0.48	-0.23	-0.7	-0.22	0.21	0.39	-0.53	-0.46	-0.37	-0.4	0.24
	7		0.55	0.25	-0.52	0.19	0.53	0.23	0.43	0.21	-0.27	-0.03	0.72	0.41
Postfire (2007)	2		0.28	0.11	0.19	0.38	-0.57	0.64	0.42	0.24	0.45	0.34	-0.43	0.51
	4		0.38	-0.69	0.31	0.44	0.18	-0.24	0.37	-0.65	0.39	0.3	0.31	-0.3
	7		0.52	0.45	0.61	-0.32	0.07	-0.22	0.41	0.31	0.37	-0.68	-0.02	-0.37
	Standard deviation		37.34	13.44	9.08	7.56	5.40	2.40	2.18	0.78	0.55	0.47	0.33	0.17
	Variance (%)		79.70	10.33	4.71	3.27	1.66	0.33	78.90	10.09	5.02	3.63	1.86	0.50

components. We know that the first and the last components are likely to be rejected. The first due to its highest variance, representing classes other than burned areas. The last due to its near-zero variance, capturing mainly noise. Hence, we focus on some of the higher order components, though, ignoring the last ones.

The mean separabilities for the components subset of our interest (meaning components 2, 3 and 4) are summarised in Table 12. The overall best PCA version for these components is the *uncentered-unscaled* one. Even in cases where centered data present relatively higher mean separabilities (in Table 12, 0.907 in case D over 0.902 in A for set 2a), we need to consider that a centered PCA redistributes greater amounts of the original variance–that is including unchanged patterns–among the higher order components.

Conclusions

The statistical evaluation shows that centering and scaling, prior to the application of SVD, operate on

Table 12 Mean separabilities for higher order principal components

Data set			Principal Components	Uncentered Unscaled (A)	Uncentered Scaled (B)	Centered Unscaled (C)	Centered Scaled (D)
MODIS	Unitemporal	1a	2, 3	0.763	0.774	0.655	0.743
	Bitemporal	2a	2, 3, 4	0.902	0.776	0.808	0.907
Landsat5 TM	Unitemporal	1b	2, 3, 4	1.439	1.352	1.438	1.288
	Bitemporal	2b		1.351	1.078	1.431	1.176
	Means			1.008	0.607	0.962	0.930

Alexandris *et al. Open Geospatial Data, Software and Standards* (2017) 2:21

Page 20 of 26

the input multi-dimensional matrix generally in a non-destructive way. If performed, centering modifies the way that data clusters are intercepted by the transformed axes. Effectively projecting spectral information related to unchanged patterns in higher order components. This works rather against the spectral enhancement of burned area clusters. Scaling smooths out fine variations existing in the original data. The latter may neutralise minor to moderate–but potentially useful details.

Within the framework of burned area mapping, the spectral separability estimations between burned and major land cover samples, point to the uncentered-unscaled SVD-based PCA version as the most suitable one. The uncentered-scaled version is rather expectedly not useful as it appears to have random effects. The centered-unscaled and centered-scaled versions should be tested. Yet, we generally discourage the use of scaling the original data if it is important to retain fine details after the transformations.

Since SVD is not optimised for class separability, centering or not centering the input data matrix, should be examined carefully. Even small improvements might be significant in further analysing the transformed data.

Endnotes

[1] eigenvector decomposition

[2] Distributed by the Land Processes Distributed Active Archive Center (LP DAAC), located at USGS/EROS, Sioux Falls, SD. http://lpdaac.usgs.gov

[3] Local Granule ID: `MOD09GQ.A2007242.h19v05.005.2007244231200.hdf`

[4] Local Granule ID: `MOD09GQK.A2006239.h19v05.004.2006241155630`

[5] Available from the U.S. Geological Survey, http://www.usgs.gov.

[6] Scene ID:`LT51830332007248MOR00`

[7] Scene ID: `LT51830332003237MTI01`

[8] Landsat Processing Details, "USGS - Landsat Missions," https://landsat.usgs.gov/landsat-processing-details (accessed April 16, 2017)

[9] Driven by the sample size restriction in GRASS-GIS' *i.smap* module, an implementation of the *SMAP* algorithm [18] to perform supervised image classification
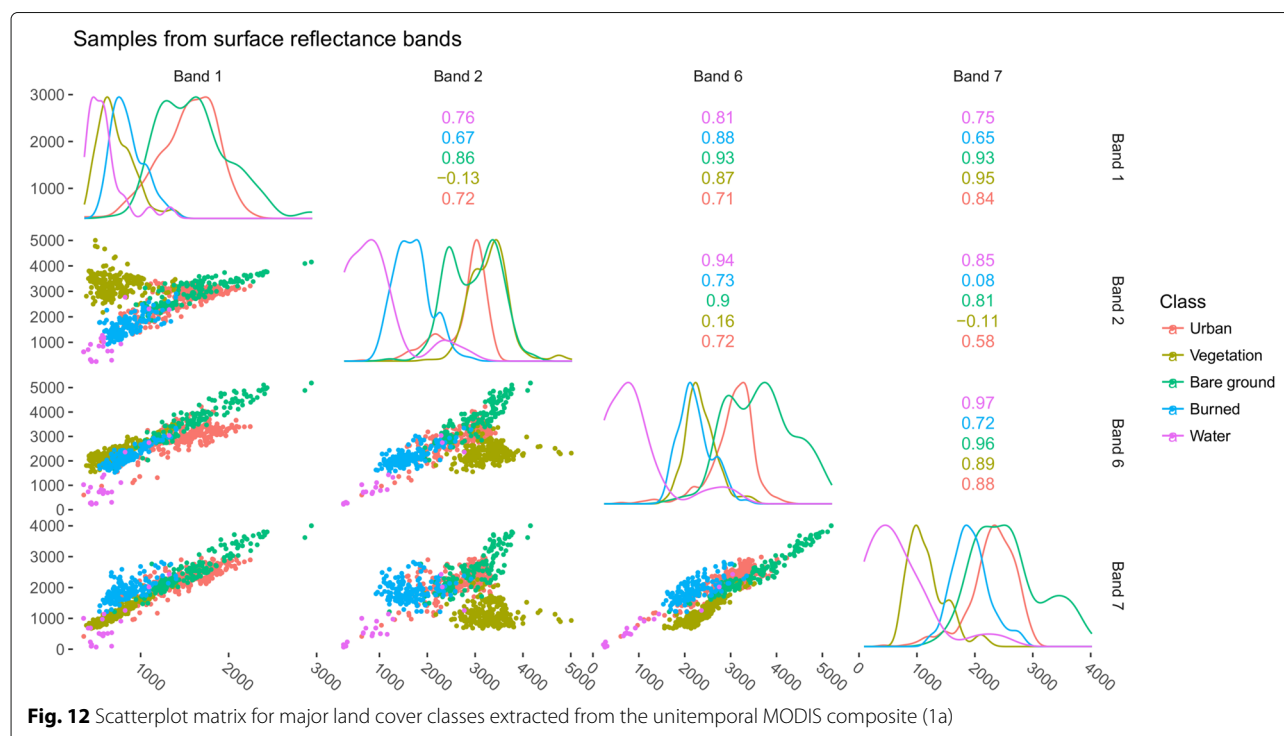
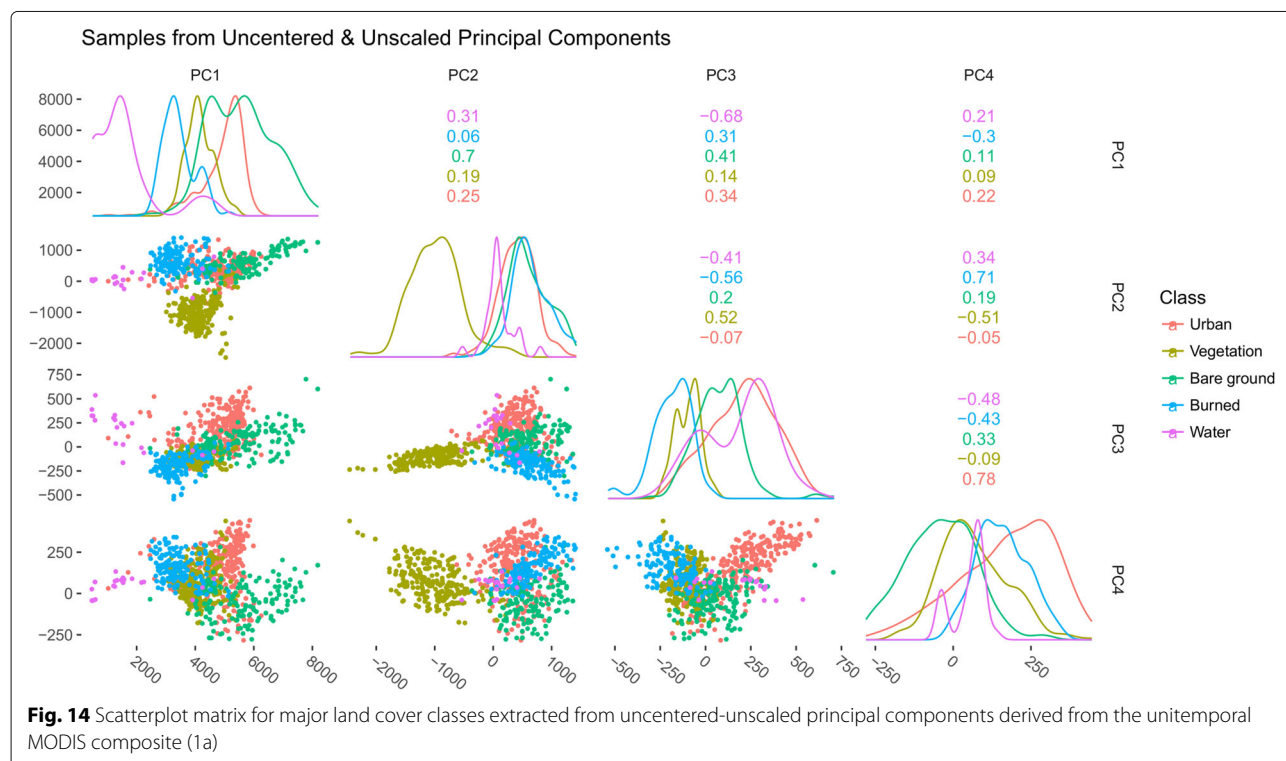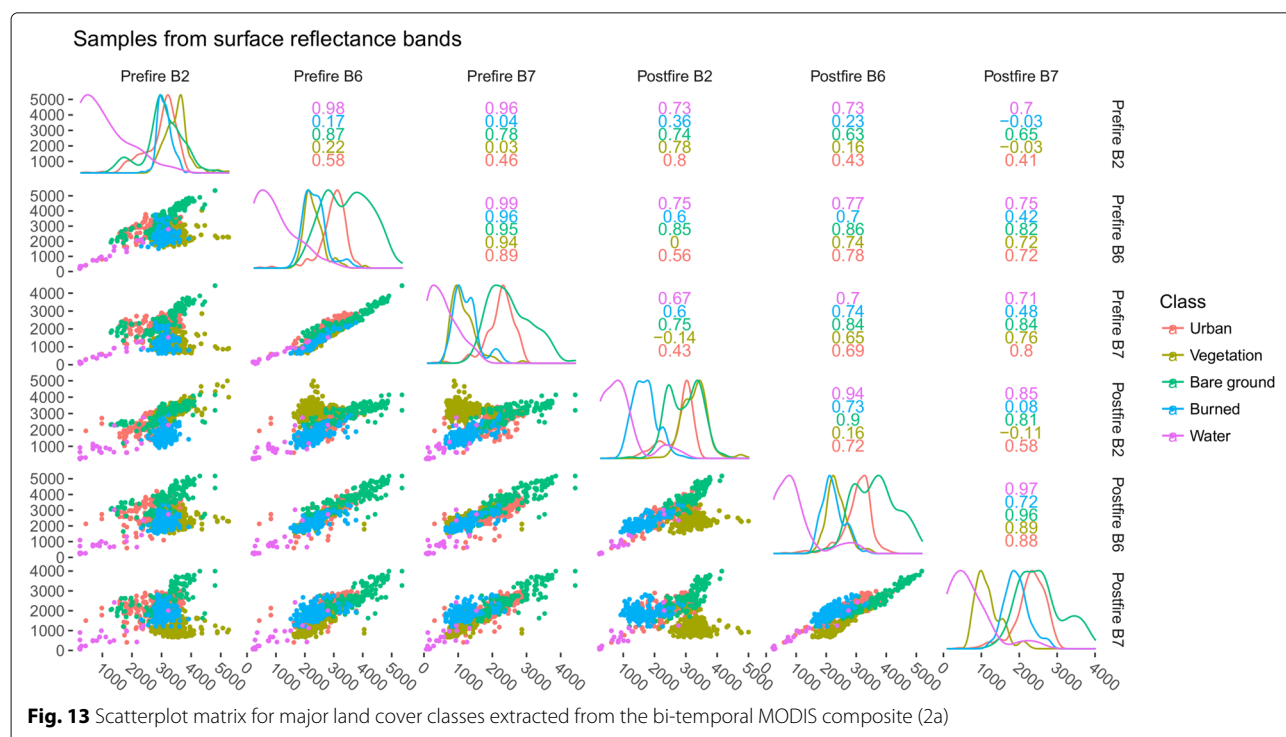[10] We use the term "class" in place "group" as used originally in the MRPP test
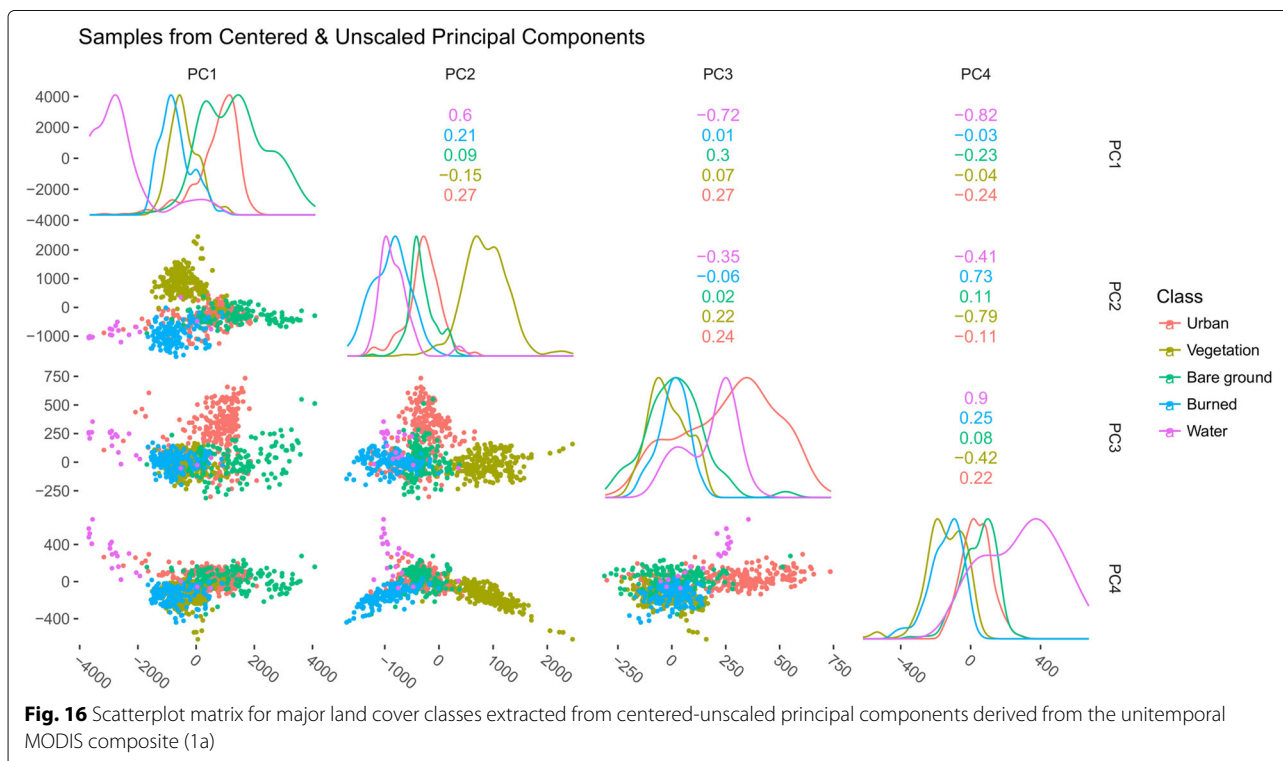
[11] *here actually singular vectors*

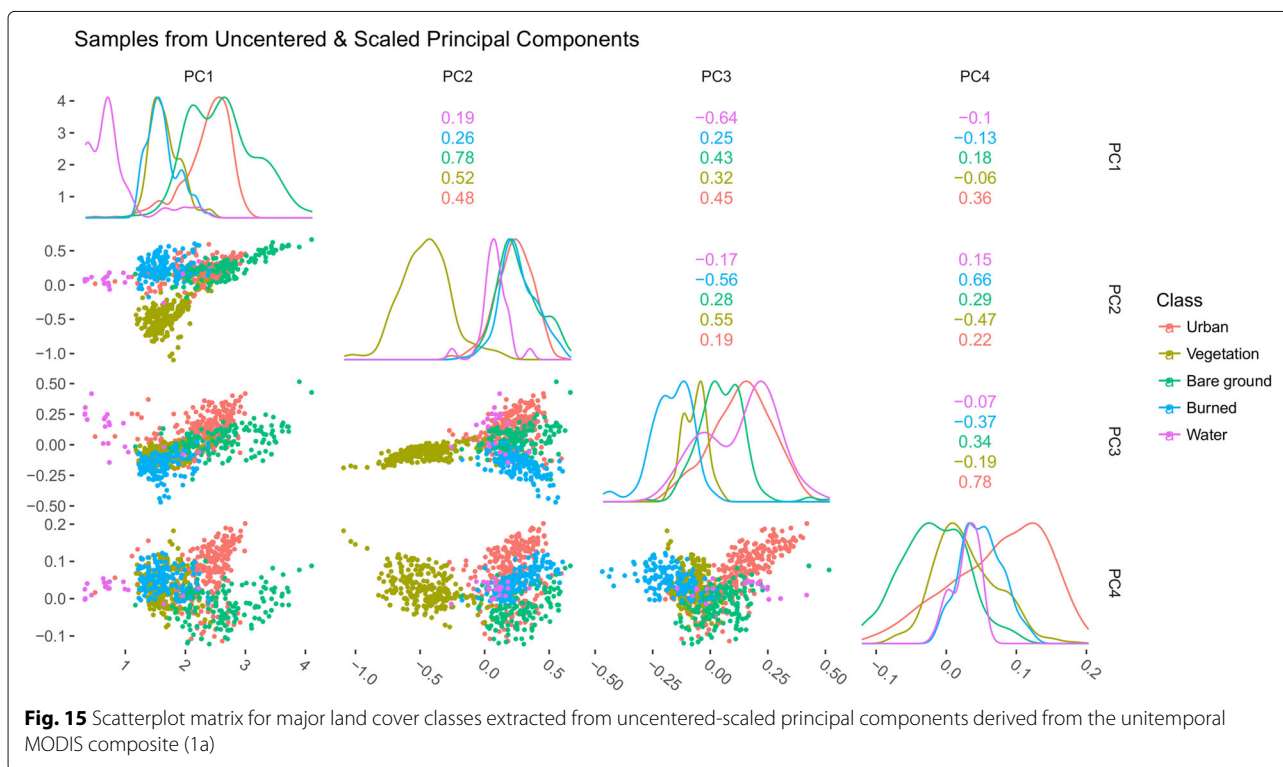[12] *vectors* can be seen as *loadings* or *weighting coefficients* which determine the direction of the principal components

[13] *here actually singular values which are square roots of non-zero eigenvalues*

[14] *eigen values* represent the variance of the original data contained in the principal components

Fig. 12 Scatterplot matrix for major land cover classes extracted from the unitemporal MODIS composite (1a)

Fig. 13 Scatterplot matrix for major land cover classes extracted from the bi-temporal MODIS composite (2a)

Fig. 14 Scatterplot matrix for major land cover classes extracted from uncentered-unscaled principal components derived from the unitemporal MODIS composite (1a)

Alexandris *et al. Open Geospatial Data, Software and Standards*   (2017) 2:21

Page 22 of 26

Fig. 15 Scatterplot matrix for major land cover classes extracted from uncentered-scaled principal components derived from the unitemporal MODIS composite (1a)

Fig. 16 Scatterplot matrix for major land cover classes extracted from centered-unscaled principal components derived from the unitemporal MODIS composite (1a)

Alexandris *et al. Open Geospatial Data, Software and Standards* (2017) 2:21

Page 23 of 26

Fig. 17 Scatterplot matrix for major land cover classes extracted from centered-scaled principal components derived from the unitemporal MODIS composite (1a)

Fig. 18 Scatterplot matrix for major land cover classes extracted from uncentered-unscaled principal components derived from the bitemporal MODIS composite (2a)

Alexandris *et al. Open Geospatial Data, Software and Standards* (2017) 2:21

Page 24 of 26

Fig. 19 Scatterplot matrix for major land cover classes extracted from uncentered-scaled principal components derived from the bitemporal MODIS composite (2a)

Fig. 20 Scatterplot matrix for major land cover classes extracted from centered-unscaled principal components derived from the bitemporal MODIS composite (2a)

Alexandris *et al. Open Geospatial Data, Software and Standards* (2017) 2:21

Page 25 of 26

Fig. 21 Scatterplot matrix for major land cover classes extracted from centered-scaled principal components derived from the bitemporal MODIS composite (2a)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Independent Researcher, Ebringen, Germany. [2]Department of Environmental and Natural Resources Management, University of Patras, Agrinio, Greece. [3]Institute of Environmental Studies, Kurukshetra University, Kurukshetra, India.

References

1. Alexandris N, Gupta S, Koutsias N. Remote sensing of burned areas via PCA. Part 1: centering, scaling and EVD vs SVD. Open Geospatial Data, Software and Standards. 2017. doi:10.1186/s40965-017-0028-1.
2. Jolliffe IT. Principal Component Analysis, 2nd edn. Springer; 2002. 28 illustrations. http://www.springer.com/statistics/statistical+theory+and+methods/book/978-0-387-95442-4.
3. Lu D, Mausel P, Brondizio E, Moran E. Change detection techniques. Int J Remote Sensing. 2003;25(12):2365. doi:10.1080/0143116031000139863.
4. Richards J, Milne A. Mapping fire burns and vegetation regeneration using principal components analysis. In: 1983 International Geoscience and Remote Sensing Symposium(IGARSS'83). San Francisco; 1983.
5. Cadima J, Jolliffe I. On relationships between uncentred and column-centred principal component analysis. Pak J Stat. 2009;25(4): 473–503.
6. Roy D, Lewis P, Justice C. Burned area mapping using multi-temporal moderate spatial resolution data - a bi-directional reflectance model-based expectation approach. Remote Sensing Environ. 2002;83: 263–86.
7. Roy D, Landmann T. Characterizing the surface heterogeneity of fire effects using multi-temporal reflective wavelength data. Int J Remote Sensing. 2005;26(19):4197–218.
8. GRASS DT. Geographic Resources Analysis Support System (GRASS GIS) Software. Open Source Geospatial Foundation, 2008. Open Source Geospatial Foundation. http://grass.osgeo.org. Accessed 28 June 2017.
9. QGIS DT. Quantum GIS Geographic Information System. Open Source Geospatial Foundation, 2009. Open Source Geospatial Foundation. http://qgis.osgeo.org. Accessed 28 June 2017.
10. Warmerdam F. FWTools: Open Source GIS Binary Kit for Windows and Linux. http://fwtools.maptools.org/. Accessed 28 June 2017.
11. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2010. R Foundation for Statistical Computing. ISBN 3-900051-07-0. http://www.R-project.org. Accessed 28 June 2017.
12. Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. Vegan: Community Ecology Package. 2010. R package version 1.17-5. http://CRAN.R-project.org/package=vegan. Accessed 28 June 2017.
13. Bossard M, Feranec J, Otahel J, Steenmans C. CORINE land cover technical guide – Addendum 2000. European Environment Agency, Kongens Nytorv 6, DK–1050 Copenhagen K, Denmark: EEA; 2000.
14. Mielke PWJ. The application of multivariate permutation methods based on distance functions in the earth sciences. Earth Science Rev. 1991;31: 55–71. doi:10.1016/0012-8252(91)90042-E.
15. Sickle JV. Using mean similarity dendrograms to evaluate classifications. J Agric Biol Environ Stat. 1997;2(4):370–88.
16. Richards J, Jia X. Remote Sensing Digital Image Analysis. An Introduction. Third, Revised and Enlarged Edition, 3rd edn: Springer; 1999, p. 363. Hard cover. ISBN 3-540-64860-7.

Alexandris *et al. Open Geospatial Data, Software and Standards*   (2017) 2:21

Page 26 of 26

17. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2016. R Foundation for Statistical Computing. https://www.R-project.org/. Accessed 28 June 2017.

18. Bouman CA, Shapiro M. A multiscale random field model for bayesian image segmentation. IEEE Trans Image Process. 1994;3(2):162–77. doi:10.1109/83.277898.